

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 00:52:06 ; Search time 73.54 Seconds  
(without alignments)  
85.089 Million cell updates/sec

Title: US-09-589-870-2  
Perfect score: 183  
Sequence: 1 MRKIVAAIVSLTTSITL.....IDAAKAGVNGNPLDAVQ 183

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 268485 seqs, 34193795 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq.36:\*

1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	183	7	P60625
2	183	100.0	183	10	P93530
3	183	100.0	183	14	R44491
4	183	100.0	183	18	W29306
5	183	100.0	183	19	W59216
6	183	100.0	183	20	Y17868
7	183	100.0	183	21	Y44701
8	183	100.0	186	21	Y60512
9	182	99.5	184	21	Y60513
10	180	87.4	161	21	Y60515
11	160	87.4	435	15	R56483
12	159	86.9	160	10	P93531

13	159	86.9	162	21	Y60514	Streptomyces avidi
14	159	86.9	163	9	P80160	Biosynthetic prote
15	152	83.1	182	8	P70492	Streptavidin sequ
16	149	81.4	183	21	Y84020	Amino acid sequenc
17	140	76.5	415	15	R56484	ScFv PRAS109 and p
18	134	73.2	159	18	W29309	Streptavidin prote
19	132	72.1	159	18	W29310	Streptavidin prote
20	128	69.9	168	21	Y44700	Potato proteinase
21	127	69.4	128	14	R34722	Core streptavidin.
22	127	69.4	128	17	W04211	Streptomyces avidi
23	127	69.4	128	17	W29308	Recombinant Core-s
24	127	69.4	270	20	Y28928	S. avidinii protein
25	127	69.4	673	17	W04208	Streptavidin/Lucif
26	123	67.2	140	17	W04210	Mutant streptomyce
27	123	67.2	685	17	W04209	Streptavidin prote
28	121	66.1	122	21	Y60516	Streptomyces avidi
29	121	66.1	122	21	Y60517	Streptomyces avidi
30	119	65.0	159	18	W29312	Streptavidin prote
31	114	62.3	159	18	W29311	Streptavidin prote
32	114	62.3	159	18	W29313	Streptavidin prote
33	114	62.3	159	18	W29314	Streptavidin prote
34	112	61.2	183	19	W59217	S. avidinii strept
35	112	61.2	183	19	W59218	S. avidinii strept
36	88	48.1	186	21	Y84021	Amino acid sequenc
37	88	48.1	186	21	Y84022	Amino acid sequenc
38	74	40.4	159	18	W29320	Streptavidin prote
39	69	37.7	402	15	R56485	ScFv PRAS110 and p
40	64	35.0	159	18	W29315	Streptavidin prote
41	64	35.0	159	18	W29316	Streptavidin prote
42	64	35.0	159	18	W29317	Streptavidin prote
43	64	35.0	159	18	W29318	Streptavidin prote
44	64	35.0	159	18	W29319	Streptavidin prote
45	40	21.9	40	8	P70491	N-terminal sequenc

#### ALIGNMENTS

RESULT	1
P60625	
ID	P60625 standard; Protein; 183 AA.
XX	
AC	P60625;
XX	
DT	13-AUG-1991 (first entry)
XX	
DE	Sequence of a streptavidin-like polypeptide encoded by SA307.
XX	
KW	Antibiotic; biotin binding affinity; fusion protein.
XX	
OS	Streptomyces.
XX	
PN	W08602077-A.
XX	
PD	10-APR-1986.
XX	
PF	01-OCT-1985; 85WO-0001901.
XX	
PR	02-OCT-1984; 84US-0656873.
XX	
PA	(MEAD/) MEADE H M.
XX	
PI	Meade HM, Garvin JL, Biogen NV;
XX	
DR	WPI: 1986-106643/16.
XX	
XX	N-PSDB; N60626.
PT	DNA sequences and hybrid DNA sequences - encoding
XX	streptavidin-like polypeptide, also joined to another protein,
XX	e.g. tissue plasminogen activator
PS	Disclosure; Fig. 2; 54pp; English.
XX	



PT positive bacterial protein.  
 XX  
 PS Disclosure: Fig 1b; 54pp; English.  
 CC Tetrameric biologically active streptavidin is produced by secretion  
 CC from *Bacillus subtilis* transformed with a plasmid encoding the  
 CC sequence.  
 CC  
 CC  
 SQ Sequence 183 AA;  
 Query Match 100.0%; Score 183; DB 14; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-173;  
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKRIIVAAIAVSLITVYITRASAADPSKDSKAQVSAEAGITWTWNLGSTRFTVAGAD 60  
 DB 1 mrlivvaalavslitvysltasaaadpskdskaqvsaeeagltgwnqlgstftvtagad 60  
 QY 61 GALTGYESAIVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATWMSGOY 120  
 DB 61 galgtgyesaivgnaesryvltgrydsapatdgsctalgwtvwmknnyrnhsatwsggy 120  
 QY 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDPFTYKPPSAASIDAAKKAGVNNGNPLDA 180  
 DB 121 vggaeairntqwlitsgtteanawksltvghdftkyvkpsaasidaakkagvnngnplda 180  
 QY 181 VQO 183  
 DB 181 vqg 183  
 RESULT 4  
 ID W29306 standard; Protein: 183 AA.  
 AC W29306;  
 XX  
 DT 27-APR-1998 (first entry)  
 XX  
 DE Wild-type streptavidin protein.  
 XX  
 KW Streptavidin: biotin; anti-interference reagent; detection; mutein;  
 KM avidin; non-specific binding.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..24  
 FT Protein /label= signal  
 FT 25..183  
 XX  
 DE19637718-A1.  
 PN  
 XX  
 PD 02-OCT-1997.  
 XX  
 PF 16-SEP-1996; 96DE-1037718.  
 XX  
 PR 01-APR-1996; 96DE-1013053.  
 XX  
 XX (BOE ) BOEHRINGER MANNHEIM GMBH.  
 PA Brandstetter H, Deger A, Engh R, Kopeckzi E, Mueller R;  
 XX Schmitt U;  
 PI  
 XX  
 DR WPI: 1997-482043/45.  
 DR N-PSDB; T73193.  
 XX  
 PT Streptavidin and avidin muteins with reduced binding affinity for  
 PT biotin - useful for reducing interference from nonspecific binding  
 PT in assays  
 XX  
 PS Disclosure: Page 17-18; 26pp; German.

XX This sequence represents a streptavidin which is used in a novel method  
 CC of reducing interference from non-specific binding in assays. Muteins  
 CC constructed from a core streptavidin or avidin sequence are selected that  
 CC differ from the native polypeptide by at least one amino acid and have a  
 CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable  
 CC polypeptide may be present as a polymeric conjugate, e.g. with another  
 CC polypeptide or protein, especially bovine serum albumin. These muteins  
 CC are used as anti-interference reagents for reducing and/or avoiding  
 CC nonspecific interactions in a process for detecting an analyte. In  
 CC particular, they are used in assays where the streptavidin/avidin-biotin  
 CC specific binding pair is involved for qualitative and/or quantitative  
 CC determination of an analyte in a test sample, e.g. a heterogeneous  
 CC immunoassay or a hybridisation assay. Despite having a lower binding  
 CC affinity for biotin, the muteins have high immunological cross-reactivity  
 CC with native streptavidin and avidin.  
 CC  
 CC  
 SQ Sequence 183 AA;  
 Query Match 100.0%; Score 183; DB 18; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-173;  
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKRIIVAAIAVSLITVYITRASAADPSKDSKAQVSAEAGITWTWNLGSTRFTVAGAD 60  
 DB 1 mrlivvaalavslitvysltasaaadpskdskaqvsaeeagltgwnqlgstftvtagad 60  
 QY 61 GALTGYESAIVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATWMSGOY 120  
 DB 61 galgtgyesaivgnaesryvltgrydsapatdgsctalgwtvwmknnyrnhsatwsggy 120  
 QY 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDPFTYKPPSAASIDAAKKAGVNNGNPLDA 180  
 DB 121 vggaeairntqwlitsgtteanawksltvghdftkyvkpsaasidaakkagvnngnplda 180  
 QY 181 VQO 183  
 DB 181 vqg 183  
 RESULT 5  
 ID W59216 standard; Protein: 183 AA.  
 AC W59216;  
 XX  
 DT 27-AUG-1998 (first entry)  
 XX  
 DE S. avidinii streptavidin protein.  
 XX  
 KW Streptavidin: ligand; binding affinity; mutant; isolation;  
 KM purification; recover; immobilise.  
 XX  
 OS Streptomyces avidinii.  
 XX  
 PN EP835934-A2.  
 XX  
 PD 15-APR-1998.  
 XX  
 PF 09-OCT-1997; 97EP-0117504.  
 XX  
 PR 10-OCT-1996; 96DE-1041876.  
 XX  
 XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.  
 PA Skerra A, Voss S;  
 XX  
 DR WPI: 1998-218668/20.  
 DR N-PSDB; V34714.  
 XX  
 PT Streptavidin mutants with higher binding affinity for peptide  
 PT ligands - have mutation in amino acid region 44-53, used to isolate,





```
XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
XX Christeller JF, Sutherland PW, Murray C, Markwick NP, Phillip BA,
XX Malone LA, Burgess EPJ;
XX WPI: 2000-171244/15.
XX N-PSDB: 249867.
XX
XX New chimeric polypeptide and composition comprising the polypeptide
XX useful for conferring pest resistance on plants -
XX
XX Disclosure: Fig 12: 111pp; English.
XX
XX The present sequence is streptavidin, a plant-toxicous protein.
XX Recombinant vector, pARR27 expressing a chimeric polypeptide comprising
XX streptavidin mature peptide fused to the potato proteinase inhibitor-II
XX (PPI-II) signal peptide is targeted to the vacuole.
XX Transformation of plant genome with the vector can produce pest
XX resistance in plants, plant derived products and stored harvest
XX material. Pests that can be controlled include, cotton bollworm,
XX tropical army-worm, European corn-borer or red mite, tobacco horn worm,
XX loopers, rice stem borer, portia, cutworms, diamondback moth, potato
XX tuber moth, codling moth, indian meal moth, gypsy moth, argentine stem
XX weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat
XX weevils, mealworms, flour beetles, black field cricket, locusts,
XX sawflies, Western flower thrips, Hessian flies or two-spotted mite.
XX
XX Sequence 183 AA:
SQ
Query Match 100.0%; Score 183; DB 21; Length 183;
Best Local Similarity 100.0%; Pred. No. 7.6e-173;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTRIVTAGAD 60
DB 1 mKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTRIVTAGAD 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGQY 120
DB 61 galTgtYesaVgnaesRyVltgrYdsapAtDgsctAlgWtVmkNnYrNaHsAtTWsgqY 120
QY 121 VGGAERINTOMLTGSTEANAMKSTLVGHDTFTYVKPSAASIDAKKAGVNGNPIDA 180
DB 121 vGgaerIntOmLTgStEaNAMkStLVghDTfTYvKpSaASIdAKKAGvNgNPIdA 180
QY 181 VQO 183
DB 181 vqg 183
RESULT 8
Y80512
ID Y80512 standard; Protein: 186 AA.
XX
XX Y80512;
XX
XX 06-JUN-2000 (first entry)
XX
XX Streptomyces avidinII sps protein.
XX
XX Plant somatic tissue degeneration; plant essential factor; depletion;
XX viability; sps gene; plant development; plant morphology; flower;
XX fruit plant.
XX
XX Streptomyces avidinII.
XX
XX WO200007427-A2.
XX
XX 17-FEB-2000.
XX
XX 30-JUL-1999; 99WO-IL00420.
XX
XX
```

```
XX
XX 03-AUG-1998; 98IL-0125632.
XX
XX (AGRI-) AGRIC RES ORG.
XX
XX Kapulnik Y, Ginzberg I;
XX
XX WPI: 2000-195402/17.
XX N-PSDB: 291073.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
XX protein, useful for controlling plant development and morphology, such
XX as decreasing the number of flowers present to increase the number of
XX fruit -
XX
XX Examples: Page 84: 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
XX plant tissue by expressing a heterologous protein capable of binding a
XX plant essential factor (PEF), in somatic plant tissue cells, where
XX heterologous protein expression causes depletion of the PEF so the plant
XX viability is maintained, while simultaneous degeneration of the somatic
XX plant tissue is effected. This sequence represents the Streptomyces
XX avidinII sps protein as an example of a heterologous protein introduced
XX into the plants. The methods can provide for the selective and optionally
XX reversible cell degeneration in somatic plant tissue. They can be used
XX for artificially controlling plant development and morphology. They can
XX be used e.g. to decrease the number of flowers in fruit producing plants
XX so as to increase the number of fruits which reach maturity.
XX
XX Sequence 186 AA:
SQ
Query Match 100.0%; Score 183; DB 21; Length 186;
Best Local Similarity 100.0%; Pred. No. 7.7e-173;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTRIVTAGAD 60
DB 4 mKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTRIVTAGAD 63
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGQY 120
DB 64 galTgtYesaVgnaesRyVltgrYdsapAtDgsctAlgWtVmkNnYrNaHsAtTWsgqY 123
QY 121 VGGAERINTOMLTGSTEANAMKSTLVGHDTFTYVKPSAASIDAKKAGVNGNPIDA 180
DB 124 vGgaerIntOmLTgStEaNAMkStLVghDTfTYvKpSaASIdAKKAGvNgNPIdA 183
QY 181 VQO 183
DB 184 vqg 186
RESULT 9
Y80513
ID Y80513 standard; Protein: 184 AA.
XX
XX Y80513;
XX
XX 06-JUN-2000 (first entry)
XX
XX Streptomyces avidinII mst protein.
XX
XX Plant somatic tissue degeneration; plant essential factor; depletion;
XX viability; mst gene; plant development; plant morphology; flower;
XX fruit plant.
XX
XX Streptomyces avidinII.
XX
XX WO200007427-A2.
XX
XX 17-FEB-2000.
XX
XX
```





CC reversible cell degeneration in somatic plant tissue. They can be used  
CC for artificially controlling plant development and morphology. They can  
CC be used e.g. to decrease the number of flowers in fruit producing plants  
CC so as to increase the number of fruits which reach maturity.

XX Sequence 162 AA;

Query Match 86.9%; Score 159; DB 21; Length 162;

Best Local Similarity 100.0%; Pred. No. 3.1e-149; Mismatches 0; Indels 0; Gaps 0;

Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 DPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALGTGESAVGNAESRYVLTGRY 84  
DB 4 dpskdskaqysaaegitgtwqngstflvtagaagaltgtyesavnaesryvltgry 63  
QY 85 DSAPATDGGSTALGWTVMKNNYRNAHSATWMSGQYVGGAEARINTQWLLTSGTTEANAW 144  
DB 64 dsapatdgsqstalgwvawknynrhaasatwsgqyvggaearintqwtllsgtteanaw 123  
QY 145 KSTLVGHDTFTKVPKPSAASIDAKKAGVNNGNPLDAVQ 183  
DB 124 kstlvghdtftkvkpsaasidaakkagvnngnpldavq 162

RESULT 14

ID P80160 standard; protein; 163 AA.

AC P80160;

DT 13-OCT-1990 (first entry)

DE Biosynthetic protein with streptavidin trailer.

XX Biosynthetic Antibody Binding site (BABS): variable domain;  
KW anti-CEA monoclonal antibody; streptavidin.

XX Synthetic.

XX W08809344-A.

PN 01-DEC-1988.

PD 19-MAY-1988; 88WO-US01737.

PF 21-MAY-1987; 87US-0052800.

PR (CREA-) CREATIVE BIOMOLECULES INC.

PI Huston JS, Oppermann H;

DR WPI; 1988-353928/49.

DR N-PSDB; N80186.

XX Recombinant multifunctional protein -  
PT having an antibody binding site and a sequence for biological activity,  
PI ion sequestering or binding to a solid support.

XX Disclosure; 115pp; English.

XX Multi-functional biosynthetic protein comprising single  
CC chain BABS and streptavidin protein trailer linked via a spacer

CC See also N80171-N80192.

XX Sequence 163 AA;

Query Match 86.9%; Score 159; DB 9; Length 163;

Best Local Similarity 100.0%; Pred. No. 3.2e-149; Mismatches 0; Indels 0; Gaps 0;

QY 25 DPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALGTGESAVGNAESRYVLTGRY 84  
DB 5 dpskdskaqysaaegitgtwqngstflvtagaagaltgtyesavnaesryvltgry 64  
QY 85 DSAPATDGGSTALGWTVMKNNYRNAHSATWMSGQYVGGAEARINTQWLLTSGTTEANAW 144  
DB 65 dsapatdgsqstalgwvawknynrhaasatwsgqyvggaearintqwtllsgtteanaw 124  
QY 145 KSTLVGHDTFTKVPKPSAASIDAKKAGVNNGNPLDAVQ 183  
DB 125 kstlvghdtftkvkpsaasidaakkagvnngnpldavq 163

RESULT 15

ID P70492 standard; protein; 182 AA.

AC P70492;

DT 06-MAR-1991 (first entry)

DE Streptavidin sequence.

XX Streptavidin; N-terminal; fusion gene; fusion protein;

XX Streptomyces avidinii.

PN W08705026-A.

PD 27-AUG-1987.

PF 24-FEB-1987; 87WO-US00397.

PR 24-FEB-1986; 86US-0833324.

PA (UYCO-) COLUMBIA UNIV N Y.

PA (UYNV-) UNIV NEW YORK.

PI Cantor CR, Axel R, Garana C;

DR WPI; 1987-250198/35.

DR N-PSDB; N70810.

XX DNA encoding streptavidin - obtd. by restriction endo-nuclease  
PT digestion of chromosomal DNA of Streptomyces avidinii

XX Disclosure; Figure 3; 54pp; English.

XX The sequence is that of streptavidin from *S. avidinii*. Streptavidin  
CC may be expressed from a fusion gene comprising its coding gene and DNA  
CC encoding a target protein of interest, where the streptavidin has  
CC binding sites for biotin or deriv. The streptavidin has 4 binding  
CC sites free for biotin, and is produced free of biotin  
CC contamination. Improved streptavidins may also be produced by  
CC site-directed mutagenesis. The fused gene may be used to produce  
CC labelled, chemically-modified proteins in vivo, and to isolate  
CC proteins when only the sequence of the gene is known.

XX Sequence 182 AA;

Query Match 83.1%; Score 152; DB 8; Length 182;

Best Local Similarity 100.0%; Pred. No. 2.8e-142; Mismatches 0; Indels 0; Gaps 0;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 AOVSAEAGITGTWYNQLGSTFIYTAGADGALGTGESAVGNAESRYVLTGRYDAPATD 91  
DB 31 aovsaeeagltgtwqngstflvtagaagaltgtyesavnaesryvltgrydsapad 90  
QY 92 GSGTALGWTVMKNNYRNAHSATWMSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGH 151  
DB 91 gsgtalgwvawknynrhaasatwsgqyvggaearintqwtllsgtteanawkstlvgh 150

Oy 152 DFTTKVPSASIDAAKKGVNNGNPLDAVQ 183  
| | | | | | | | | | | | | | | | | | | | | |  
Db 151 dtftkvkpsaasidaakkagvnmgnpldavq 182

Search completed: April 6, 2001, 01:25:30  
Job time: 2004 sec



**THIS PAGE BLANK (USPTO)**

9b_pat1:AS93649	183.00	3357.64	7.3e-179	638	U036391	Streptomyces avidinii	9
9b_pat1:AS9365	183.00	3357.64	7.3e-179	638	U036391	Streptomyces avidinii	9
9b_pat1:AS9152	160.00	2928.51	5.8e-155	1356	A93649	Sequence 1 from Patent E	9
9b_pat1:AS9152	160.00	2928.51	5.8e-155	1356	A93655	Sequence 4 from Patent	8
9b_pat1:AR082490	160.00	2928.51	5.8e-155	1356	A93152	Sequence 4 from Patent	8
9b_pat1:AS93838	159.00	2916.15	2.8e-154	498	AR082490	Sequence 4 from Patent	8
9b_pat1:U09059	159.00	2916.15	2.8e-154	498	A93838	Sequence 7 from Patent V	9
9b_pat1:A000743	159.00	2916.04	2.9e-154	507	I09059	Sequence 38 from Patent	8
9b_pat1:A000744	159.00	2916.04	2.9e-154	507	A00743	s.avidinii synthetic gene	9
9b_pat1:A020698	159.00	2915.82	3.0e-154	525	A00744	s.avidinii synthetic gene	9
9b_pat1:U15647	159.00	2915.82	3.0e-154	525	A20698	FXA-Stripa DNA sequence.	9
9b_pat1:U01349	152.00	2785.43	5.4e-147	638	I15647	Sequence 2 from patent U	9
9b_pat1:AS93567	140.00	2559.61	2.0e-134	1296	I01349	Sequence 2 from Patent U	9
9b_pat1:A93154	140.00	2559.61	2.0e-134	1296	A93567	Sequence 6 from Patent	8
9b_pat1:AR082491	140.00	2559.61	2.0e-134	1296	A93154	Sequence 6 from Patent	8
9b_sy:AF618REP	128.00	2345.28	1.8e-122	99	AR082491	Sequence 6 from Patent	8
9b_sy:AF618REP	128.00	2345.28	1.8e-122	99	X65082	Synthetic gene for strep	9
9b_sy:ASV14583	128.00	2338.31	4.3e-122	1254	AF184718	Synthetic construct s	9
9b_sy:SCFV18200	128.00	2330.73	1.4e-121	4354	ASV14583	Artificial sequence; v	9
9b_pat1:AS9663	127.00	2330.72	1.4e-121	4363	Y18290	Expression vector pSTE	9
9b_pat1:U14742	127.00	2327.05	1.8e-121	384	A93663	Sequence 15 from Patent	8
9b_pat1:AS9569	127.00	2319.81	1.9e-121	387	I14742	Sequence 1 from patent U	9
9b_pat1:AS93156	127.00	2319.83	4.6e-121	1257	A93569	Sequence 8 from Patent	8
9b_pat1:AR082492	127.00	2319.83	4.6e-121	1257	A93156	Sequence 8 from Patent	8
em_pat1:E010169	127.00	2316.95	6.7e-121	2019	AR082492	Sequence 8 from patent	8
9b_pat1:E010170	123.00	2230.25	3.5e-117	625	E010169	DNA encoding streptococ	9
9b_pat1:E010170	123.00	2230.25	3.5e-117	625	S78777	streptavidin v1 (Streptoc	9
9b_pat1:S78782	123.00	2243.01	8.8e-117	2055	E010170	DNA encoding mutated st	9
9b_pat1:AS9389	59.00	1068.90	2.2e-51	625	S78782	streptavidin v2 (Streptid	9
9b_pat1:U148879	31.00	563.53	3.1e-23	95	A13369	Oligonucleotide 1 (95 me	9
9b_pat1:U148879	31.00	563.53	3.1e-23	95	A13369	Oligonucleotide probe (4	9
9b_pat1:U148879	31.00	563.53	3.1e-23	95	A13369	Oligonucleotide probe (4	9
9b_pat1:U148879	21.00	380.63	4.8e-13	72	I19879	Sequence 78 from patent U	9
9b_pat1:U148879	19.00	344.43	5.0e-11	64	I19880	Sequence 79 from patent U	9
9b_pat1:AS93655	14.00	252.64	6.5e-06	59	A93655	Sequence 7 from Patent E	9
9b_pat1:AS94480	13.00	232.80	8.2e-05	74	A92480	Sequence 11 from Patent E	9
9b_pat1:AS93653	12.00	217.23	0.0006	46	A93653	Sequence 5 from Patent E	9
9b_pat1:AS93654	12.00	217.23	0.0006	46	A93654	Sequence 6 from Patent E	9
9b_pat1:AS93657	11.00	186.39	0.0008	49	A93657	Sequence 11 from Patent E	9
9b_pat1:AS93657	9.00	152.71	0.6609	40	A93657	Sequence 9 from Patent E	9

gb_pat1:A936556	+	9.00	162.71	0.6609	40	A93658	Sequence 10 from Paten
gb_pat1:A93660	+	9.00	162.71	0.6609	40	A93662	Sequence 12 from Paten
gb_pat2:ECR076	+	9.00	136.03	20.23	3200	X71365	E.chrysanthemi gene
gb_ov:GGY18681	-	9.00	135.81	20.82	3319	Y18681	Gallus gallus genom
gb_pat1:AE005021	+	9.00	129.02	49.75	10128	AE005021	Halobacterium sp.
gb_pat1:AE000094	-	9.00	128.99	49.89	10166	AE000094	Rhizobium sp. NGR

200 TCGACCTTCATCGTGACCGGGGGCCGACGGCGCCCTGACCGGAACCTA 24

67 rgiuserAlaValGlyValAsnAlaGluSerArgTyrValLeuThrGlyArgT 84  
|||||  
250 CGAGTCGCGCGTGGCCACGCCGCAAGAGCCCTACGTCTTACCGGTCGT 299  
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPThr 100  
|||||  
300 ACAGACAGCGCCCGCGCCACCGGAGCGGACCGCCCTCGGTGGAGC 349  
101 ValAlaThrPlysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrPse 117  
|||||  
350 GTGGCTCGGAAGATACCTACCGCAAGCCCTCCGCGCACCAAGTGGAG 399  
117 rgiYglNtYrValGlyGlyValaGluAlaArgIleAsnThrGlnTyrPLeu 134  
|||||  
400 CGGCGACGTACGTGGCGGCGCCGCAAGCGAGATCAACACCCAGTGGCTGC 449  
134 eutHrSerGlyThrThrGluAlaAsnAlaThrPlysSerThrLeuValGly 150  
|||||  
450 TGACCTCCGCGCACCGAGCCGCTGGAAGTCCACGCTGGTCGCGC 499  
151 HisAspThrThrThrLysValLysProSerAlaAlaSerIleAspAlaAl 167  
|||||  
500 CACGACACCTTCACCAAGGTAAGCGCTCCGCGCTCCATCGACGCGCGC 549  
167 alySLysAlaGlyValaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183  
|||||  
550 GAAGAAGCGCGCGCTCAACAACGCAACCCGCTCGACGCGCTTCAGCAG 598

seq\_name: gb\_pat1:A93649

seq\_documentation\_block: 638 bp DNA PAT 22-JAN-2000  
LOCUS A93649  
DEFINITION Sequence 1 from Patent EP0799890.  
ACCESSION A93649  
VERSION A93649.1 GI:6741838  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 638)  
AUTHORS Mueller, R.D. and Deger, A.D.  
TITLE Recombinant inactive core streptavidin mutants  
JOURNAL Patent: EP 0799890-A 08-OCT-1997;  
BOEHRINGER MANNHEIM GMBH (DE)

FEATURES  
Source Location/Qualifiers  
1..638

sig\_peptide /organism="unidentified"  
/db\_xref="taxon:32644"  
50..121  
CDS 50..601  
/note="unnamed protein product"

/codon\_start=1  
/protein\_id="CAB69587.1"  
/db\_xref="GI:6741839"  
/translation="MKRIVAAIAVSLTTSITASADPSKDSKAQVSAABAGITGT  
MYNGLGSFTITAGADGALGCTYSAGVANSARYLGRYDASAPATQSSGTALGWTVA  
MKNYRNHNSHTTWSGQYVGAEARINTQMLTSGTTAANKSTLVGHDTFTYKPS  
AASIDAAKAGVNGNPLDAVQ"

mat\_peptide 115 a 244 c 193 g 86 t  
BASE COUNT 122..598  
ORIGIN

alignment\_scores:  
Quality: 183.00 Length: 183  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-589-870-2 x A93649 ..

Align seg 1/1 to: A93649 from: 1 to: 638

1 MetArgPylIleValValAlaAlaIleAlaValSerLeuThrThrValse 17  
|||||  
50 ATGGCGAAGATCGTGTGGACCATGCGCTTCCCTTGACACAGGCTTC 99  
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGln 34  
|||||  
100 GATTACGGCCAGCGCTTCGGCAGACCCCTCCAAAGACTCCAAAGGCCAGG 149  
34 aIleSerAlaAlaGluAlaGlyIleThrGlyThrTyrPThrAsnGlnLeuGly 50  
|||||  
150 TCTCGCGCGCGGAGCGCGGATCACCGCACCTGGTAAACACAGCTTCGCG 199  
51 SerThrPheIleValThrAlaGlyValaAspGlyValaLeuThrGlyThrTy 67  
|||||  
200 TCGACCTTCATCTGACCGCGGCGCGCGACGCGCTTGACCGGACCTTA 249  
67 rgiuserAlaValGlyValaAsnAlaGluSerArgTyrValLeuThrGlyArgT 84  
|||||  
250 CGAGTCGCGCGTGGCCACGCCGCAAGCCGCTACGTCTTACCGGTCGT 299  
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPThr 100  
|||||  
300 ACGACACGCGCCCGCGCCACCGAGCGGACGCGCACCGCCCTCGGTGGAGC 349  
101 ValAlaThrPlysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrPse 117  
|||||  
350 GTGGCTCGGAAGATACCTACCGCAAGCCCTCCGCGCACCAAGTGGAG 399  
117 rgiYglNtYrValGlyGlyValaGluAlaArgIleAsnThrGlnTyrPLeu 134  
|||||  
400 CGGCGACGTACGTGGCGGCGCCGAGGCGAGATCAACACCCAGTGGCTGC 449  
134 eutHrSerGlyThrThrGluAlaAsnAlaThrPlysSerThrLeuValGly 150  
|||||  
450 TGACCTCCGCGCACCGACCGAGCGCAACGCTGGAAGTCCACGCTGGTCGCGC 499  
151 HisAspThrThrThrLysValLysProSerAlaAlaSerIleAspAlaAl 167  
|||||  
500 CACGACACCTTCACCAAGGTAAGCCCTCCGCGCTCCATCGACGCGGC 549  
167 alySLysAlaGlyValaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183  
|||||  
550 GAAGAAGCGCGCGCTCAACAACGCAACCCGCTCGACGCGCTTCAGCAG 598

seq\_name: gb\_pat1:A39565

seq\_documentation\_block: 1356 bp DNA PAT 05-MAR-1997  
LOCUS A39565  
DEFINITION Sequence 4 from Patent WO9415644.  
ACCESSION A39565  
VERSION A39565.1 GI:2295847  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1356)  
AUTHORS Epenetos, A.A., Spooner, R.A. and Deonarain, M.  
TITLE COMPOUNDS FOR TARGETING  
JOURNAL Patent: WO 9415644-A 4 21-JUL-1994;  
IMP CANCER RES TECH (GB)  
COMMENT Other publication GB 2289679 951129.  
FEATURES  
Source Location/Qualifiers  
1..1356  
/organism="unidentified"  
/db\_xref="taxon:32644"  
40..1347  
/note="unnamed protein product"

CDS

/codon\_start=1  
/protein\_id="CAA02449.1"  
/db\_xref="GI:2295848"  
/translation="MKRYLLPFAAGILLLAOPAMAQVLOQPGAEIVKPGASVKLSC  
KASGYTPSYMHVWKQRPGRGLEMIGRIDPNSGDTKYNKFKSKATITVDKPSSTAY



MOLESITSEDSAVVYICARYDYSSYFDYWGQTTVTYSSGGGGGGGGGGGSOAV  
VTQESALTPGCVTLTKRSSTGAVTTSNVTNMYOEKPDHLEFTGLIGSTNNRAGVP  
ARFSGSLIDKALITGAOTDEALYFCALYNSHMYWGGGKTLVLGLLEAPAAPA  
DPSKDSKAOVSAEAGITGTWYNOLGSPFIYAGADALGTYESVGNABERYVLTG  
RTDSAPATDGGSTALGWTVAAMKNYRNAAHSATTWGQYVGAEBARINTQWLTSCTTE  
ANAMKSTLVGHDTFTKVKRPSAASIDAARKAGVNNGNPLDAVQO"

BASE COUNT 332 a 357 c 361 g 306 t  
ORIGIN

alignment\_scores:  
Quality: 160.00 Length: 160  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-589-870-2 x A39565 ..

Align seg 1/1 to: A39565 from: 1 to: 1356

```
24 AAlaasproSerLysaspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCCGCTCCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGAAGCTGG 914

40 YLleThrGlyThrTrpTyraSngInLeuGlySerThrPheIleValThra 57
|||||
915 TATACCTGGCACCTGGTGTATACCAACTGGGCTCCACTTCTCTTGACCG 964

57 laGlyAlaAspGlyAlaLeuThrGlyThrTyrgLuserAlaValAlaGlysn 73
|||||
965 CTGGTGGGAGAGGAGCTGTGACTGGACCTGAGATCTGCGGTTGGTGAAC 1014

74 AAlaGuserArGtyrValLeuThrGlyArGtyrAspSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACGGCGCTTATGACTCTGCACCTGCCAC 1064

90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
|||||
1065 CGATGGCTCTGTACCGCTCTGGGCTGGACTGTGGCTTTGGAAAAACAAC 1114

107 yTrArGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValAlaGly 123
|||||
1115 ATCGTAATGCCACACAGCGCCACTGACGTGCTGGCCAAATACGTTGGCG 1164

124 AAlaGluAlaArGlyIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
1165 GCTGAGGCTGCTATCAACACTCAGTGGCTGTAAACATCCGGCAGCTACCG 1214

140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAG 1264

157 aLysProSerAlaAlaSerIleAspAlaAlaLysAlaGlyValAlaSn 173
|||||
1265 TTAAAGCCTTGTGCTAGCATGATTGCTGCCAAGAAAGAGCGGTAAC 1314

174 AsnGlyAsnProLeuAspAlaValAlaGln 183
|||||
1315 AACGTTAACCTCTAGACGCTGTTACAGCA 1344
```

seq\_name: gb\_pat1:A93152

seq\_documentation\_block: 1356 bp DNA PAT 22-JAN-2000

LOCUS A93152

DEFINITION Sequence 4 from Patent EP0815872.

ACCESSION A93152

VERSION A93152.1 GI:6741540

KEYWORDS

SOURCE unidentiflied.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1356)  
AUTHORS Epenetos,A.A. and Deonaraia,M.

TITLE Compounds for targeting  
JOURNAL Patent: EP 0815872-A 07-JAN-1998;  
IMP CANCER RES TECH (GB)

FEATURES  
source Location/Qualifiers  
1..1356  
/organism="unidentified"  
/db\_xref="taxon:32644"

CDS  
40..1347  
/note="unnamed protein product"

/protein\_id="CAB69540.1"  
/db\_xref="GI:6741541"

/translation="MKYLPTRAAGLLLAQAPAMAQVQLDQAPAEIVPGASVKLSC  
KASGYFTSWMHWKORPGRLEWIRDPNSGGTKYERKSRATLTVDPSSRYAY  
VTQESALTPGCVTLTKRSSTGAVTTSNVTNMYOEKPDHLEFTGLIGSTNNRAGVP  
ARFSGSLIDKALITGAOTDEALYFCALYNSHMYWGGGKTLVLGLLEAPAAPA  
DPSKDSKAOVSAEAGITGTWYNOLGSPFIYAGADALGTYESVGNABERYVLTG  
RTDSAPATDGGSTALGWTVAAMKNYRNAAHSATTWGQYVGAEBARINTQWLTSCTTE  
ANAMKSTLVGHDTFTKVKRPSAASIDAARKAGVNNGNPLDAVQO"

BASE COUNT 332 a 357 c 361 g 306 t  
ORIGIN

alignment\_scores:  
Quality: 160.00 Length: 160  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-589-870-2 x A93152 ..

Align seg 1/1 to: A93152 from: 1 to: 1356

```
24 AAlaasproSerLysaspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCCGCTCCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGAAGCTGG 914

40 YLleThrGlyThrTrpTyraSngInLeuGlySerThrPheIleValThra 57
|||||
915 TATACCTGGCACCTGGTGTATACCAACTGGGCTCCACTTCTCTTGACCG 964

57 laGlyAlaAspGlyAlaLeuThrGlyThrTyrgLuserAlaValAlaGlysn 73
|||||
965 CTGGTGGGAGAGGAGCTGTGACTGGACCTGAGATCTGCGGTTGGTGAAC 1014

74 AAlaGuserArGtyrValLeuThrGlyArGtyrAspSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACGGCGCTTATGACTCTGCACCTGCCAC 1064

90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
|||||
1065 CGATGGCTCTGTACCGCTCTGGGCTGGACTGTGGCTTTGGAAAAACAAC 1114

107 yTrArGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValAlaGly 123
|||||
1115 ATCGTAATGCCACACAGCGCCACTGACGTGCTGGCCAAATACGTTGGCG 1164

124 AAlaGluAlaArGlyIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
1165 GCTGAGGCTGCTATCAACACTCAGTGGCTGTAAACATCCGGCAGCTACCG 1214

140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAG 1264

157 aLysProSerAlaAlaSerIleAspAlaAlaLysAlaGlyValAlaSn 173
|||||
1265 TTAAAGCCTTGTGCTAGCATGATTGCTGCCAAGAAAGAGCGGTAAC 1314

174 AsnGlyAsnProLeuAspAlaValAlaGln 183
|||||
1315 AACGTTAACCTCTAGACGCTGTTACAGCA 1344
```

```

seq_name: gb_pat1:AR082490
seq_documentation_block:
LOCUS AR082490 1356 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 4 from patent US 5973116.
ACCESSION AR082490
VERSION AR082490.1 GI:10009216
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1356)
AUTHORS Epenetos,A.,Antoniou, Spooner,R.,Anthony and Deonarain,M.
TITLE Compounds for targeting
JOURNAL Patent: US 5973116-A 4/26-OCT-1999;
FEATURES
source 1..1356
/organism="unknown"
BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN

```

```

alignment_scores:
Quality: 160.00 Length: 160
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment\_block:

US-09-589-870-2 x AR082490 ..

Align seg 1/1 to: AR082490 from: 1 to: 1356

```

24 AlaasproserlysaspserysAlaGlnValSerAlaAlaGlnAlaG1 40
|||||
865 GCAGACCCGTCAGGACCTCCAAAGCTCAGGTTCTGCAGCCGAAGCTGG 914
40 YilethG1YThrTPYrAsnGlnLeuGlySerThrPheIleValThra 57
|||||
915 TATCACCTGGACCTGCTATTAACCAACTGGGGTGCACCTTCATTGTGACCG 964
57 laGlyAlaaspg1yAlaLeuThrGlyThrTyrg1uSerAlaValaGlyAsn 73
|||||
965 CTGGTGGGAGCGAGCTGACTGCGACCTACGAATCTGCGGTGTGAAC 1014
74 AlaGluSerArgTyValLeuThrGlyArgTyAspSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCGCTACGTACTGCGCGTTATGACTCTGCACCTGCCAC 1064
90 rAspGlySerG1YThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsn 107
|||||
1065 CGATGGCTCTGGTACCCCTCTGGGCTGACTGTGGCTTGGAAAAACAAC 1114
107 yTrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyValGlyGly 123
|||||
1115 ATCGTATGCGACAGCGCCACTGCTGTGCGCAATGACGTTGGCGGT 1164
124 AlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
1165 GCTGAGGCTGTATCAACACTGAGTGGCTGTAAACATCCGCGCATACGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAAATGCGTGAATCGACACTAGTAGTCATGACACCTTTTCCAAAG 1264
157 aLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsn 173
|||||
1265 TTAAAGCTTCTGCTGCTAGCATGTATGCTGCAAGAAAGCAGCGGTAAAC 1314
174 AsnGlyAsnProLeuAspAlaValaGlnGln 183
|||||
1315 AACGTTAACCTCTAGACGCTGTTCAGCAA 1344

```

```

seq_name: gb_pat1:A93838
seq_documentation_block:
LOCUS A93838 498 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9728263.
ACCESSION A93838
VERSION A93838.1 GI:6741969
KEYWORDS
SOURCE
ORGANISM unidentified.
REFERENCE
1 (bases 1 to 498)
AUTHORS Lubitz,W. and Sleytr,U.
TITLE RECOMBINANT EXPRESSION OF S-LAYER PROTEINS
JOURNAL Patent: WO 9728263-A 07-AUG-1997;
LUBITZ WERNER (AT); SLEYTR UWE (AT)
FEATURES
source 1..498
/db_xref="taxon:32644"
BASE COUNT 120 a 137 c 131 g 110 t
ORIGIN

```

```

alignment_scores:
Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment\_block:

US-09-589-870-2 x A93838 ..

Align seg 1/1 to: A93838 from: 1 to: 498

```

25 AspProserLysAspSerLysAlaGlnValSerAlaAlaGlnAlaG1 41
|||||
7 GACCCGTCCAAAGACTCCAAAGCTCAGGTTCTGCACCCGAAGCTGGTAT 56
41 eThrG1YThrTPYrAsnGlnLeuGlySerThrPheIleValThraLag 58
|||||
57 CACTGGCACTCGTATTAACCACTGGGGTGCACCTTCTATTGTGACCCCTG 106
58 lYAlaaspg1yAlaLeuThrGlyThrTyrg1uSerAlaValaGlyAsnAla 74
|||||
107 GTGGGAGCGAGCTCTGACTGCGACCTACGAATCTGCGGTGTGAACGA 156
75 GluSerArgTyValLeuThrGlyArgTyAspSerAlaProAlaThras 91
|||||
157 GAATCCCGCTACGTACTGCGCGTTATGACTCTGCACCTGCCACCGA 206
91 pGlySerG1YThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyra 108
|||||
207 TGCTCTGGTACCCCTCTGGGCTGACTGTGGCTTGGAAAAACAACATATC 256
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyValGlyAla 124
|||||
257 GTAATGGGACAGCGCCACTGCTGTGCGCAATGACGTTGGCGGTGCT 306
125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG1ua1 141
|||||
307 GAGGCTGTATCAACACTGAGTGGCTGTAAACATCCGCGCATACCGAGC 356
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysVal 158
|||||
357 GAATGCAATGGAATCGACACTAGTAGTCATGACACCTTTTCCAAAGTTA 406
158 ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValaLysAsn 174
|||||
407 AGCCTTCTGCTGCTAGCATGTATGCTGCCAAGAAAGCAGCGGTAAACAA 456
175 GlyAsnProLeuAspAlaValaGlnGln 183
|||||
457 GGTAAACCTCTAGACGCTGTTCAGCAA 483

```

```

seq_name: gb_pat1:109059
seq_documentation_block:
LOCUS      109059           498 bp          PAT          02-DEC-1994
DEFINITION Sequence 38 from Patent WO 8809344.
ACCESSION  109059
VERSION    109059.1  GI:588242
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 498)
AUTHORS    Huston, J.S. and Oppermann, H.
TITLE      TARGETED MULTIFUNCTIONAL PROTEINS
JOURNAL    Patent: WO 8809344-A 38 01-DEC-1988;
FEATURES   Location/Qualifiers
            source          1..498
                                /organism="unknown"
BASE COUNT  117 a      137 c      133 g      111 t
ORIGIN
alignment_scores:
            Quality: 159.00      Length: 159
            Ratio: 1.000      Gaps: 0
            Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-589-870-2 x 109059      ..
Align seg 1/1 to: 109059 from: 1 to: 498
25  AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI1 41
13  GACCCGTCGAAGGACTCCAAAGCTCAGGTTTCTGCTGCCGAAGCTGGTAT 62
41  eThhGlyThrTyrTrpTyrAsnGlnLeuGlySerThrPheIleValThrIaG 58
63  CACTGGCAGCTGGTATTAACCAACTGGGGTGCACCTTTCATTGTGTACCCCTG 112
58  LysAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValAlaGlyAsnAla 74
113  GTGGGAGCGAGCTCTGACTGCGCACTACGAATCGCGGTTGGTGAACGCA 162
75  GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
163  GAATCCCGCTACGATGACTGACGCGTTATGACTGTCGACCTGCCACGCA 212
91  pGlySerGlyThrAlaLeuGlyTTPThrValAlaLTrpLysAsnAsnTyrA 108
213  TGGCTCTGGTACCGCTCTGGCTGGACTGTGCTTGGAAAAACAATATC 262
108  rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValAlaGlyAla 124
263  GTAATGGCGACAGCGCCACTACGTGGTCTGCCCAATACGTTGGCGGTCT 312
125  GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
313  GAGCGCTGATCAACACTGACGTGCTGTTAACTACCGGCACTACCGAAGC 362
141  eAsnAlaLTrpLysSerThrLeuValAlaGlyHisAspThrPheThrIysValL 158
363  GAATGCAATGGAAATCGACACTAGTAGTCATGACACCTTTACCAAGTTA 412
158  ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
413  AGCCTTCTGCTGCTAGCATTTGATCTGCGCAAGAAAGCAGCGGTAAACAC 462
175  GlyAsnProLeuAspAlaValAlaGln 183
463  GGTAAACCTTAGACGCTGTTCACGAA 489

```

```

seq_name: gb_pat1:A00743
seq_documentation_block:
LOCUS      A00743           507 bp          DNA          PAT          28-JAN-1993
DEFINITION S.avidinii synthetic gene for streptavidin.
ACCESSION  A00743
VERSION    A00743.1  GI:14605
KEYWORDS
SOURCE     Streptomyces avidinii.
ORGANISM   Streptomyces avidinii.
REFERENCE  1 (bases 1 to 507)
AUTHORS    Edwards, R.
TITLE      SYNTHETIC GENE
JOURNAL    Patent: WO 8903422-A 2 20-APR-1989;
FEATURES   Location/Qualifiers
            source          1..507
                                /organism="Streptomyces avidinii"
                                /db_xref="taxon:1895"
                                /note="synthetic gene"
                                /codon_start=1
                                /transl_table=11
                                /product="streptavidin"
                                /protein_id="CAA0084.1"
                                /db_xref="GI:14606"
                                /translation="MDPSKSKAOKVSAAEAGITGTWYNQSGSTFTVYAGADALNGTY
                                ESAVGNASREYVLTGRYDSAPATDGSGLGCTVAMKNNYRNAHSATTWSCGVGAE
                                ARINQWLLTSGTTEANAMKSTLVGHDFFTKVPASASIDAAKKAGVNNGNPLDAVQO"
CDS
BASE COUNT  125 a      137 c      130 g      115 t
ORIGIN
alignment_scores:
            Quality: 159.00      Length: 159
            Ratio: 1.000      Gaps: 0
            Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-589-870-2 x A00743      ..
Align seg 1/1 to: A00743 from: 1 to: 507
25  AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI1 41
13  GACCCGTCGAAGGACTCCAAAGCTCAGGTTTCTGCTGCCGAAGCTGGTAT 62
41  eThhGlyThrTyrTrpTyrAsnGlnLeuGlySerThrPheIleValThrIaG 58
63  CACTGGCAGCTGGTATTAACCAACTGGGGTGCACCTTTCATTGTGTACCCCTG 112
58  LysAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValAlaGlyAsnAla 74
113  GTGGGAGCGAGCTCTGACTGCGCACTACGAATCGCGGTTGGTGAACGCA 162
75  GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
163  GAATCCCGCTACGATGACTGACGCGTTATGACTGTCGACCTGCCACGCA 212
91  pGlySerGlyThrAlaLeuGlyTTPThrValAlaLTrpLysAsnAsnTyrA 108
213  TGGCTCTGGTACCGCTCTGGCTGGACTGTGCTTGGAAAAACAATATC 262
108  rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValAlaGlyAla 124
263  GTAATGGCGACAGCGCCACTACGTGGTCTGCCCAATACGTTGGCGGTCT 312
125  GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
313  GAGCGCTGATCAACACTGACGTGCTGTTAACTACCGGCACTACCGAAGC 362

```

141 aasnaIatrpIySerThrLeuValGlyHisAspThrPheThrIysValL 158  
 |||  
 363 GAATGCATGGAATGCACACTAGTACATGACACCTTACCAAGTTA 412  
 |||  
 158 yspProSerAlaIaSerIleAspAlaIaIySlyAlaGlyValaAsn 174  
 |||  
 413 AGCCTTCTGCTGCTAGCATGTGATGCGCAAGAAAGAGCGTAAACAC 462  
 |||  
 175 GlyAsnProLeuAspAlaValGlnGln 183  
 |||  
 463 GGTAAACCTCTAGACGCTGTTCAGCAA 489

seq\_name: gb\_pat1:A00744

seq\_documentation\_block:

LOCUS A00744 507 bp DNA PAT 28-JAN-1993  
 DEFINITION S.avidinii synthetic gene (reverse complement) for streptavidin.

ACCESSION A00744

VERSION A00744.1 GI:14607

KEYWORDS

SOURCE Streptomyces avidinii.

ORGANISM Streptomyces avidinii.

REFERENCE Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

AUTHORS Edwards, R.

TITLE SYNTHETIC GENE

JOURNAL Patent: WO 8903422-A 3 20-APR-1989;

British Bio-Technology Ltd

FEATURES

1.507

/db\_xref="taxon:1895"

BASE COUNT 115 a 130 c 137 g 125 t

ORIGIN

alignment\_scores:

Quality: 159.00 Length: 159

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x A00744/rev ..

Align seg 1/1 to reverse of: A00744 from: 1 to: 507

25 AspProSerIysAspSerIysAlaGlnValSerAlaIaGlyI 41  
 |||  
 495 GACCCGTCAGGACCTCCAAAGCTAGCTTCTGACGCGAAGCGGTAT 446  
 |||  
 41 eThrGlyThrTyrTrpAsnGlnLeuGlySerThrPheIleValThrIaG 58  
 |||  
 445 CACTGGCAGCTGTATTAACCAACTGGGCTGCACTTTCATTTGTGACCGCTG 396  
 |||  
 58 IyAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAla 74  
 |||  
 395 GTGGGAGGAGAGCTGCTGACCTGACCTAGCAATCTGCGGTTGGTAACCA 346  
 |||  
 75 GluSerArgTyrValIleuThrGlyArgTyrAspSerAlaProAlaThrAs 91  
 |||  
 345 GAATCCCGCTAGTACTGACTGCGGCTTATGACTGTCGACCTGCCACCGA 296  
 |||  
 91 pGlySerGlyThrAlaLeuGlyTyrThrValAlaIaTrpIysAsnAsnTyrA 108  
 |||  
 295 TGGCTCTGTGCTACCGCTGCGGCTGAGCTGTGGTGAAGAAACAATATC 246  
 |||  
 108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyIyAla 124  
 |||  
 245 GTAATGGCAGACAGCGCACTAGCTGCTGCGCAATACGTTGGCGGTCT 196

125 GluAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrIaGlyI 141  
 |||  
 195 GAGGCTCGATACACACTCAGTGGCTGTTAACCTCGGACACTACGGAAGC 146  
 |||  
 141 aasnaIatrpIySerThrLeuValGlyHisAspThrPheThrIysValL 158  
 |||  
 145 GAATGCATGGAATGCACACTAGTACATGACACCTTACCAAGTTA 96  
 |||  
 158 yspProSerAlaIaSerIleAspAlaIaIySlyAlaGlyValaAsn 174  
 |||  
 95 AGCCTTCTGCTGCTAGCATGTGATGCGCAAGAAAGAGCGTAAACAC 46  
 |||  
 175 GlyAsnProLeuAspAlaValGlnGln 183  
 |||  
 45 GGTAAACCTCTAGACGCTGTTCAGCAA 19

seq\_name: gb\_pat1:A20698

seq\_documentation\_block:

LOCUS A20698 525 bp DNA PAT 22-AUG-1994  
 DEFINITION Fxa-StrpA DNA sequence.

ACCESSION A20698

VERSION A20698.1 GI:583400

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 525)

AUTHORS CARRIER-BOUND RECOMBINANT PROTEIN, PROCESS FOR PRODUCING IT AND ITS

TITLE USE AS AN IMMUNOGEN AND VACCINE

JOURNAL Patent: WO 9113155-A 2 05-SEP-1991;

Location/Qualifiers

1.525

/db\_xref="taxon:32630"

BASE COUNT 128 a 140 c 138 g 119 t

ORIGIN

alignment\_scores:

Quality: 159.00 Length: 159

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x A20698 ..

Align seg 1/1 to: A20698 from: 1 to: 525

25 AspProSerIysAspSerIysAlaGlnValSerAlaIaGlyI 41  
 |||  
 37 GACCCGTCAGGACCTCCAAAGCTAGCTTCTGACGCGAAGCGGTAT 86  
 |||  
 41 eThrGlyThrTyrTrpAsnGlnLeuGlySerThrPheIleValThrIaG 58  
 |||  
 87 CACTGGCAGCTGTATTAACCAACTGGGCTGCACTTTCATTTGTGACCGCTG 136  
 |||  
 58 IyAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAla 74  
 |||  
 137 GTGGGAGGAGAGCTGCTGACCTGACCTAGCAATCTGCGGTTGGTAACCA 186  
 |||  
 75 GluSerArgTyrValIleuThrGlyArgTyrAspSerAlaProAlaThrAs 91  
 |||  
 187 GAATCCCGCTAGTACTGACTGCGGCTTATGACTGTCGACCTGCCACCGA 236  
 |||  
 91 pGlySerGlyThrAlaLeuGlyTyrThrValAlaIaTrpIysAsnAsnTyrA 108  
 |||  
 237 TGGCTCTGTGCTACCGCTGCGGCTGAGCTGTGGTGAAGAAACAATATC 286  
 |||  
 108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyIyAla 124  
 |||  
 287 GTAATGGCAGACAGCGCACTAGCTGCTGCGCAATACGTTGGCGGTCT 336

```
125 GUAlAAlaArgTlleAsnThrGlnTrpLeuLeuThrSerGlyThrTrpGluAl 141
|||||
337 GAGGCTGTATCAACACTCAAGTGGCTTTAACTCCGGCACTACCGAAGC 386
|||||
141 AAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
|||||
387 GAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAACTTA 436
|||||
158 ySPROSerAlaAlaSerTlleAspAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
437 AGCCTTCTGCTGCTAGCATGATGCTGCCAAGAAACGAGCGGTAAACAC 486
|||||
175 GlyAsnProLeuAspAlaValGlnGln 183
|||||
487 GGTAACTCTTAGAGCTGTTCAGCAA 513
|||||

seq_name: gb_pat1:115647

seq_documentation_block:
LOCUS 115647 525 bp DNA PAT 02-APR-1996
DEFINITION Sequence 2 from patent US 5470573.
ACCESSION 115647
VERSION 115647.1 GI:1250555
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 525)
AUTHORS Lubitz,W. and Szostak,M.P.
TITLE Immunoens comprising the non-lytic membrane spanning domain of
JOURNAL Patent: US 5470573-A 2 28-NOV-1995;
FEATURES
source 1..525
location/Qualifiers
BASE COUNT 128 a 140 c 138 g 119 t
ORIGIN

alignment_scores:
Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 115647 ..

Align seg 1/1 to: 115647 from: 1 to: 525

25 AsPProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyTl 41
|||||
37 GACCCGTCGAAGGACTCCAAAGCTCAGTTCTGCACCGCAAGCTGGTAT 86
|||||
41 eThrGlyThrTrpTrpAsnGlnLeuGlySerThrPheLleValThrAlaG 58
|||||
87 CACTGGCACTGGTATACCAACTGGGGTGCATTTCAATTGTGACCCCTG 136
|||||
58 LylAAspGlyAlaLeuThrGlyThrTyrgLysSerAlaValAlaGlyAsnAla 74
|||||
137 GTGGCGAGGAGGCTGCTACGTGGCAGCTACGAATGTGGGTTGGTAAAGCA 186
|||||
75 GluSerAlaGlyValAlaLeuThrGlyAlaGlyTyrAspSerAlaProAlaThrAs 91
|||||
187 GAATCCCGCTACGACTGACTGGCCGCTTATGACTCTGACCTGCCACGA 236
|||||
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyrAla 108
|||||
237 TGGCTGTGGTACCGCTGTGGCTGGACTGTGGTGAATAAACACTATTC 286
|||||
108 rGAAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyAlaAla 124
|||||
287 GTATATGGCAACAGCGCACTACGTGTCTGGCCAAATACGTTGGCGGTGCT 336
|||||
```

```
125 GUAlAAlaArgTlleAsnThrGlnTrpLeuLeuThrSerGlyThrTrpGluAl 141
|||||
337 GAGGCTGTATCAACACTCAAGTGGCTTTAACTCCGGCACTACCGAAGC 386
|||||
141 AAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
|||||
387 GAATGCATGGAATCGACACTAGTAGTCATGACACCTTTACCAAACTTA 436
|||||
158 ySPROSerAlaAlaSerTlleAspAlaAlaLysLysAlaGlyValAsnAsn 174
|||||
437 AGCCTTCTGCTGCTAGCATGATGCTGCCAAGAAACGAGCGGTAAACAC 486
|||||
175 GlyAsnProLeuAspAlaValGlnGln 183
|||||
487 GGTAACTCTTAGAGCTGTTCAGCAA 513
|||||

seq_name: gb_pat1:101349

seq_documentation_block:
LOCUS 101349 638 bp ss-DNA PAT 21-MAY-1993
DEFINITION Sequence 2 from Patent US 4839293.
ACCESSION 101349
VERSION 101349.1 GI:270135
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 638)
AUTHORS Cantor,C.R., Axel,R. and Argarana,C.
TITLE DNA encoding streptavidin, streptavidin produced therefrom, fused
polypeptides which include amino acid sequences present in
JOURNAL Patent: US 4839293-A 2 13-JUN-1989;
The Trustees of Columbia University in the City of New York;
FEATURES
source 1..638
location/Qualifiers
BASE COUNT 114 a 244 c 194 g 86 t
ORIGIN

alignment_scores:
Quality: 152.00 Length: 152
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 101349 ..

Align seg 1/1 to: 101349 from: 1 to: 638

32 AlaGlnValSerAlaAlaGluAlaGlyTlleThrGlyThrTrpTyrAsnG 48
|||||
143 GCCCAAGTCTCGCGCCGCGAGCGCGGCATCACCGCACTGTGTAAACCA 192
|||||
48 nLeuGlySerThrPheLleValThrAlaGlyAlaAspLysAlaLeuThrG 65
|||||
193 GCTGGCTCGACTTCATCGTGAACGGGGGGCGGAGCGGCGCTGACCG 242
|||||
65 LylThrTyrgLysSerAlaValAlaGlyAsnAlaGlySerAlaGlyValAlaLeuThr 81
|||||
243 GAACCTACGAGTTCGGCGCTGGCAACGCCGAGACCCGCTACGCTTGACC 292
|||||
82 GlyAlaGlyAspSerAlaProAlaThrAspLysSerGlyThrAlaLeuG 98
|||||
293 GGTGTTACGACACGCCGCCGCGCACCGACGCGGACCGGCTTCGG 342
|||||
98 YTrpThrValAlaTrpLysAsnAsnTyrAlaGAsnAlaHisSerAlaThrT 115
|||||
343 TTGCAAGGTGGCGTGAATAATACCTACCGCAAGCCGCACTGCCGACCA 392
|||||
```

```
115 htrpserglylntryvalglyalaglualarglleasnthrcln 131
|||||
393 CGTGGAGCGCCAGTACGTGCGGCGCGGAGCGAGATCAACCCAG 442
|||||
132 trypleuethrserglythrthrghualaasnatrplysSerThrLe 148
|||||
443 TGCGTGTGACCTCCGGCACACCGAGCCGCTGGAAGTCCACGCT 492
|||||
148 uvalglyhisasprhrptherthlysalysproserAlaIAserIleA 165
|||||
493 GGTGGCGCACGACACCTTCACCAAGGTGAACCGTCCGCCCTCATCG 542
|||||
165 spAlaAlaIAslysaIaglyvalAsnaSnglyAsnProleuaspAlaVal 181
|||||
543 ACGGGGGAAGAGAGCGCGGCGCTCAACAGCGCAACCGCTCGACGCCGT 592
|||||
182 GlnGln 183
|||||
593 CAGCAG 598

seq_name: gb_patl:A39567
seq_documentation_block:
LOCUS A39567 1296 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 6 from Patent WO9415644.
ACCESSION A39567
VERSION A39567.1 GI:2295849
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE
1 (bases 1 to 1296)
Epenetos/A.A., Spooner,R.A. and Deonaraïn,M.
AUTHORS
COMPOUNDS FOR TARGETING
TITLE
Patent: WO 9415644-A 6 21-JUL-1994;
JOURNAL
IMP CANCER RES TECH (GB)
COMMENT
Other publication GB 2289679 951129.
FEATURES
source
location/Qualifiers
1..1296
/organism="unidentified"
/db_xref="taxon:32644"
40..1287
/feature="unnamed protein product"
/codon_start=1
/protein_id="CA02450.1"
/db_xref="GI:2295850"
/translation="MKYLLPTAAAGLLLLAQAAPMAVOYLOQPEAEIVKPGASVKLSL
KASGYTFSTYMMHWKORPRGLEWIGRIDNSGSTRYNEKFSKATLTVDKPSSTAY
MOLSLTSEDSAVYYCARYDYGSSYFDYWGQGTYYVSSGGSGGGSGGSGSOAY
VQDESALTTSPEETVLTICRSTGAVTTSNANWVQEKPDHLFTGLIGTNNRPAQVP
ARFSGSLIGDKAALITGAQTDEDAIYFCALWYSNHWYFGGTGLVIGLEAPAAFA
DPSKDSKAQVSAEAGITGTWVNOIGSTFIYTAGADGALITGYESAVGNMESRYVING
RDSAFAITDGSCTALGWTYAMKNNYRNHNSATVTSWGOYVGAEARINTQWLITSGTTE
ANAKKSTLVGHDTFTVKRPSAAS"

BASE COUNT 312 a 342 c 347 g 295 t
ORIGIN
alignment_scores:
Quality: 140.00 Length: 140
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x A39567 ..
Align seg 1/1 to: A39567 from: 1 to: 1296
24 AlaasprSerlyasprSerlyasAlaGlnValSerAlaAlaGlnAlaGln 40
|||||
865 GCAGACCGCGTCCAGAGACTCCAAAGCTCAGTTCTGCAGCGCGAGCTGG 914
|||||
40 ylierhnglythrtrptryasnGlnleuGlyserThrPheIleValThrA 57
|||||
```

```
|||||
915 TATCACTGGCACCTGTGTAACCACTGGGGTGCACCTTCATTTGACCG 964
|||||
57 IaglyAlaasprGlyAlaLeuThrGlyThrIYrgLuserAlaValaGlyAsn 73
|||||
965 CTGGTCGAGACGAGGCTCTGACTGGCAGCTACGAAATCTGCGGTGGTAAAC 1014
|||||
74 AlaGuserArGlyValleuThrGlyArGlyrAspSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGTACTACTGACTGACGCGCTGTATGACTCTGCACCTGCCAC 1064
|||||
90 rAspGlySerGlyThrAlaLeuGlyThrIYrgLuserAlaAlaIAserIleA 107
|||||
1065 CGAGGCTCTGTAACCGCTCTGGGCTGACACTGTGGTGGAAACAACACT 1114
|||||
107 YrArGAsnaAlaHisSerAlaThrThrTrpSerGlylntryValaGlyGly 123
|||||
1115 ATCGTAAATGGCGACAGCGCCACTACGTGCTGGCCAAATACGTTGGCGGT 1164
|||||
124 AlaGlnAlaArglleasnThrGlntrpLeuLeuThrSerGlyThrThrGln 140
|||||
1165 GCTGAGGCTCGTATCAACACTCAGTGGCTGTAAACATCCGCGCACTACCGA 1214
|||||
140 uAlaasnaIAsrplysSerThrleuValGlyHisasprhrptherthlyAsy 157
|||||
1215 AGCGAATCGATGAAATCGACACTAGTACATGACACACCTTTACCAAG 1264
|||||
157 alAsprSerAlaIAser 163
|||||
1265 TTAAAGCTTGTGCTGTACG 1284

seq_name: gb_patl:A93154
seq_documentation_block:
LOCUS A93154 1296 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0815872.
ACCESSION A93154
VERSION A93154.1 GI:6741542
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE
1 (bases 1 to 1296)
Epenetos/A.A. and Deonaraïn,M.
AUTHORS
COMPOUNDS for targeting
TITLE
Patent: EP 0815872-A 07-JAN-1998;
JOURNAL
IMP CANCER RES TECH (GB)
COMMENT
Location/Qualifiers
1..1296
/organism="unidentified"
/db_xref="taxon:32644"
40..1287
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAB69541.1"
/db_xref="GI:6741543"
/translation="MKYLLPTAAAGLLLLAQAAPMAVOYLOQPEAEIVKPGASVKLSL
KASGYTFSTYMMHWKORPRGLEWIGRIDNSGSTRYNEKFSKATLTVDKPSSTAY
MOLSLTSEDSAVYYCARYDYGSSYFDYWGQGTYYVSSGGSGGGSGGSGSOAY
VQDESALTTSPEETVLTICRSTGAVTTSNANWVQEKPDHLFTGLIGTNNRPAQVP
ARFSGSLIGDKAALITGAQTDEDAIYFCALWYSNHWYFGGTGLVIGLEAPAAFA
DPSKDSKAQVSAEAGITGTWVNOIGSTFIYTAGADGALITGYESAVGNMESRYVING
RDSAFAITDGSCTALGWTYAMKNNYRNHNSATVTSWGOYVGAEARINTQWLITSGTTE
ANAKKSTLVGHDTFTVKRPSAAS"

BASE COUNT 312 a 342 c 347 g 295 t
ORIGIN
alignment_scores:
Quality: 140.00 Length: 140
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment\_block:  
US-09-589-870-2 x A93154 ..

Align seg 1/1 to: A93154 from: 1 to: 1296

```
24 ALaasProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCCGCTCCAAAGAGCTCAAGAGCTCAGGTTTCTGCAGCCGAAGCTGG 914
40 YIleThrgLysThrTrpTyraSngInLeuGlySerThrPheIleValTrpA 57
|||||
915 TATCACTGGCACCTGGTATTAACAACCTGGGGCTGCACCTTTCATTGGACCG 964
57 lAgLYAlaAspGlyValAlaLeuThrGlyThrTyrgLysSerAlaValGlyasn 73
|||||
965 CTGGTGGCGAGAGAGAGCTGTGACTGGCAGCTTACGAATCTGCGGTTGGTAAc 1014
74 AlagLysSerArgTyraValLeuThrGlyYargTyraSeraLaproAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACTGGCCGCTTATGACTCTGCACCTGCCAC 1064
90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
|||||
1065 CGATGGCTCTGTGTAACCGCTGTGGCTGGAGCTGTGGAAAAACAACt 1114
107 YrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyraValGlyGly 123
|||||
1115 ATCGTAATGCCACAGCCGCTACGTGGCTGTGGCAATACGTTGGCGGT 1164
124 AlagLualAargIleasnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
1165 GCTGAGGCTCGTATCAACACACAGTGCGCTGTTAACATCCGGCACACACGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGGAATGCACACAGTAGTGCATGACACCTTACCAAG 1264
157 aLYsProSerAlaAlaSer 163
|||||
1265 TTAAGCTTCTGCTGCTAGC 1284
```

seq\_name: gb\_pat1:AR082491

seq\_documentation\_block:

```
LOCUS AR082491 1296 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 6 from patent US 5973116.
ACCESSION AR082491
VERSION AR082491.1 GI:10009217
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Epenetos,A,Antoniou, Spooner,R,Anthony and DeonaraIn,M.
TITLE Compounds for targeting
JOURNAL Patent: US 5973116-A 6 26-OCT-1999;
FEATURES
source
1..1296
/organism="Unknown"
BASE COUNT 312 a 342 c 347 g 295 t
ORIGIN
```

alignment\_scores:

Quality: 140.00 Length: 140  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x AR082491 ..

Align seg 1/1 to: AR082491 from: 1 to: 1296

```
24 ALaasProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCCGCTCCAAAGAGCTCAAGAGCTCAGGTTTCTGCAGCCGAAGCTGG 914
40 YIleThrgLysThrTrpTyraSngInLeuGlySerThrPheIleValTrpA 57
|||||
915 TATCACTGGCACCTGGTATTAACAACCTGGGGCTGCACCTTTCATTGGACCG 964
57 lAgLYAlaAspGlyValAlaLeuThrGlyThrTyrgLysSerAlaValGlyasn 73
|||||
965 CTGGTGGCGAGAGAGAGCTGTGACTGGCAGCTTACGAATCTGCGGTTGGTAAc 1014
74 AlagLysSerArgTyraValLeuThrGlyYargTyraSeraLaproAlaTh 90
|||||
1015 GCAGAAATCCCGCTACGTACTGACTGGCCGCTTATGACTCTGCACCTGCCAC 1064
90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
|||||
1065 CGATGGCTCTGTGTAACCGCTGTGGCTGGAGCTGTGGAAAAACAACt 1114
107 YrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyraValGlyGly 123
|||||
1115 ATCGTAATGCCACAGCCGCTACGTGGCTGTGGCAATACGTTGGCGGT 1164
124 AlagLualAargIleasnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
1165 GCTGAGGCTCGTATCAACACACAGTGCGCTGTTAACATCCGGCACACACGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysV 157
|||||
1215 AGCGAATGCATGGAATGCACACAGTAGTGCATGACACCTTACCAAG 1264
157 aLYsProSerAlaAlaSer 163
|||||
1265 TTAAGCTTCTGCTGCTAGC 1284
```







```

|||||
1 ATGCCAGATGTCGTGCACACCATCGCGCTTCCCTACACAGCTCTC 50
17 rtleThraSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
51 GATPACGGCAGCGCTCGGCGAGACCCCTCCAAGGACTCGAAGGCCAG 100
34 a1SerAlaGluAlaGlyIleThrGlyThrTrpTyraAsnGlnLeuGly 50
|||||
101 TCTCGCGCGCGAGCGGCATCATCCGCGACCTGGTACCAACGCTCGGC 150
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
131 TCGACCTTCATGTCGACCGCGGCGCCGACGCGCCCTGACCGGAACTTA 200
67 rglUserAlaValGlyAsnAlaGluSerArgTyValLeuThrGlyArgT 84
201 CGAGTCGGCGCGCGCAACGCGCGAGCGGCTACGTCCTGACCGGTCGT 250
84 y1AspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
251 ACGACAGCGCGCGCGCACCGCGAGCGGCGACCGCCCTCGGTTGGAGC 300
101 ValAlaTrpLysAsnAsnTyraArgAsnAlaHisSerAlaThrThrTrpSe 117
301 GTGGCGCTGGAAGAATACTACCGCAACGCCCACTCCGCGACACGCTGAG 350
117 rglYglntYrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
351 CGGCCAGTACGTCGGCGCGCGCGAGCGGCGATCAACACCCAGTGGCTGC 400
134 eutHrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
401 TGACCTCGGCGCACCGACCGAGCGCAAGCCTGGAACTCCACGCTGCTGGC 450
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
451 CACGACACCTTCACCAAGGTGAAGCGCTCGCGGCTCCATGAGAGCGGC 500
167 aLysLysAlaGlyValaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
501 GAAGAGGCGCGCGCTCAACAACGCGCAACCCGCTCGACGCGCTTCAGCAG 549
seq_name: /cgrn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT.291073
seq_documentation_block:
ID 291073 standard; DNA; 563 BP.
XX
AC 291073;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomycetes avidin1 sps gene.
XX
KW plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; sps gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomycetes avidin1.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PE 30-JUL-1999; 99WO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
DR WPI; 2000-195402/17.

```

```

DR P-PSDB; Y80512.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 84; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomycetes
CC avidin1 sps gene as an example of the heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 563 BP; 107 A; 213 C; 169 G; 74 T; 0 other;

alignment_scores:
Quality: 183.00 Length: 183
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 291073 ..

Align seg 1/1 to: 291073 from: 1 to: 563

1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSe 17
|||||
10 ATGCCAGATGTCGTGTCACACCATCGCGCTTCCCTACACAGCTCTC 59
17 rtleThraSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
|||||
60 GATPACGGCAGCGCTCGGCGAGACCCCTCCAAGGACTCGAAGGCCAGG 109
34 a1SerAlaGluAlaGlyIleThrGlyThrTrpTyraAsnGlnLeuGly 50
|||||
110 TCTCGCGCGCGAGCGGCATCATCCGCGACCTGGTACCAACGCTCGGC 159
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
160 TCGACCTTCATGTCGACCGCGGCGCCGACGCGCCCTGACCGGAACTTA 209
67 rglUserAlaValGlyAsnAlaGluSerArgTyValLeuThrGlyArgT 84
|||||
210 CGAGTCGGCGCGCGCAACGCGCGGCTACGTCCTGACCGGTCGT 259
84 y1AspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
|||||
260 ACGACAGCGCGCGCGCACCGCGAGCGGCGACCGCCCTCGGTTGGAGC 309
101 ValAlaTrpLysAsnAsnTyraArgAsnAlaHisSerAlaThrThrTrpSe 117
|||||
310 GTGGCGCTGGAAGAATACTACCGCAACGCCCACTCCGCGACACGCTGAG 359
117 rglYglntYrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
360 CGGCCAGTACGTCGGCGCGCGAGCGGCGATCAACACCCAGTGGCTGC 409
134 eutHrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
410 TGACCTCGGCGCACCGACCGAGCGCAAGCCTGGAACTCCACGCTGCTGGC 459
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
|||||

```



```

FT   sig_peptide      50..121
FT   /*tag= b
FT   mat_peptide     122..598
FT   /*tag= c
FT   /product= Streptavidin
XX
XX   EP835934-A2.
XX   15-APR-1998.
XX
XX   09-OCT-1997; 97EP-0117504.
XX
XX   10-OCT-1996; 96DE-1041876.
XX
XX   (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX   Skeria A, Voss S;
XX   PI
XX   WPI: 1998-218868/20.
XX   P-PSDB; W59216.
XX
XX   Streptavidin mutants with higher binding affinity for peptide
XX   ligands - have mutation in amino acid region 44-53, used to isolate,
XX   purify or determine fusion proteins including these ligands
XX
XX   Disclosure; Page -: 21pp; German.

```

```

XX   This sequence encodes a wild-type streptavidin protein isolated from
XX   Streptomyces avidinii. This sequence is used to produce mutants which
XX   are used in a method to assay the binding affinity of streptavidin
XX   mutants. These mutants have a mutation within the amino acid (aa) region
XX   44-53 of the wild-type protein show a higher binding affinity than the
XX   wild-type for peptide ligands that include the sequence of formula
XX   R1P-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly,
XX   or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can
XX   be used to isolate, purify and determine proteins or to determine/recover
XX   substances that contain streptavidin-binding groups. Such compounds may
XX   also be used to immobilise fusions on microtitre plates, microbeads or
XX   sensor chips.
XX   NOTE: This sequence does not appear in the specification but is used to
XX   make the mutant streptavidin sequence represented in V34715 and V34716.
XX
XX   Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

```

```

Alignment_scores:
Quality: 183.00      Length: 183
Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment\_block:

US-09-589-870-2 x V34714 ..

Align seg 1/1 to: V34714 from: 1 to: 638

```

1 MetArgLysIleValValAlaIleAlaValSerIleuThrThrValSe 17
|||||
50 ATGCCCAAGATGCTGTGCACGCCATCGCCGTTCCCTGCACCGAGCTCC 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34
|||||
100 GATTACGGCCAGCGCTTGGCGCAGACCCCTCCAGAGACTCGAAGGCCGACG 149
|||||
34 alSerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnleugly 50
|||||
150 TCTCGCGCGCGAGCGCGCATCACCGCAGACTGTGTACACACAGCTCGCGC 199
|||||
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
200 TCGACCTTCATCGTGACCGCGCGCGCCGACGCGCCGTGACCGGAACCTA 249
|||||
67 rGluSerAlaValAlaGluAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||

```

```

250 CGAGTCGGCGCTGCGCAACGCCAGAGCCGCTACGTCGTGACCGGTCGT 299
84 yraspserAlaProAlaThrAspGlySerGlyThrAlaLeuIlyTrpThr 100
|||||
300 ACGACAGCGCCCGCGCCAGCGAGCGAGCGACCGCCCTCGGTTGGAGC 349
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
|||||
350 GTGGCCTGGAAGATRACATCCGCAACGCCCACTCCGCGACACAGCTGGAG 399
117 rGlyGlnTyrValGlyGlyAlaGluAlaArgTlleasnThrGlnTrpLeuL 134
|||||
400 CGGCCAGTACGTGCGGGCGCGAGCGAGATCAACACCGAGTGGCTGC 449
134 euThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrIleuValGly 150
|||||
450 TGACCTCCGCGCACACCGAGCGCAACGCTGGAAGTCCACGCTGGTGGCGC 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaL 167
|||||
500 CAGGACACCTTCACCAAGGTGAAGCGGTCCGCGCTCCATCGACGCGGC 549
167 alYstLysAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 GAAGAAAGCGCGCGTCAACACGCAACGCCCGCTCGACGCGCTTCAGCAG 598

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA2000.DAT:Z49867

seq\_documentation\_block:

ID Z49867 standard; cDNA; 638 BP.

XX Z49867;

XX 25-APR-2000 (first entry)

XX Streptavidin gene for recombinant pART27 vector.

XX potato proteinase inhibitor-II; PPI-II; streptavidin; worm;

KW insect; plant-noxious protein; pest resistance; moth; insect; weevil;

KW grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;

XX insecticidal; ss.

XX OS Synthetic.

XX Key Location/Qualifiers

XX CDS 50..601

XX /\*tag= a

XX /product= "Streptavidin protein"

XX sig\_peptide 50..121

XX /\*tag= b

XX mat\_peptide 122..598

XX /\*tag= c

XX WO200004049-A1.

XX 27-JAN-2000.

XX 15-JUL-1999; 99WO-N200110.

XX 15-JUL-1998; 98NZ-0331002.

XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.

XX Christeller JT, Sutherland PW, Murray C, Markwick NP, Phillip BA;

XX Malone LA, Burgess EPJ;

XX WPI: 2000-171244/15.

XX P-PSDB; Y44701.

XX New chimeric polypeptide and composition comprising the polypeptide

XX useful for conferring pest resistance on plants

XX Disclosure; Fig 12; 11pp; English.

XX The present sequence encodes streptavidin, a plant-noxious protein. This  
 CC is used in the preparation of a recombinant vector, pART27 designed to  
 CC express a chimeric polypeptide comprising streptavidin mature peptide  
 CC fused to the potato protease inhibitor-II (PI-II) signal peptide. The  
 CC binary vector is targeted to the vacuole by PI-II signal sequence.  
 CC Transformation of plant genome with the vector can produce pest  
 CC resistance in plants, plant derived products and stored harvest  
 CC material. Pests that can be controlled include: cotton bollworm,  
 CC tropical army-worm, European corn-borer or red mite, tobacco horn worm,  
 CC loopers, rice stem borer, porina, cutworms, diamondback moth, potato  
 CC tuber moth, codling moth, Indian meal moth, gypsy moth, argentine stem  
 CC weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat  
 CC weevils, mealworms, flour beetles, black field cricket, locusts,  
 CC sawflies, Western flower thrips, Hessian flies or two-spotted mite.

XX Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other:

## alignment\_scores:

Quality: 183.00 Length: 183  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-589-870-2 x 249867 ..

Align seg 1/1 to: 249867 from: 1 to: 638

```

1 MetArgLysIleValValAlaIleAlaValSerLeuThrValSe 17
  |||||
50 ATGCCGACATGCTGTTGACCATCGCGTTCCCTGACCAACGGTCTC 99
  |||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGln 34
  |||||
100 GATTACGGCAGCGCTTCGACACCCCTCCAAAGACTCGAAAGCCGACAG 149
  |||||
34 aISerAlaIleAlaIleValIleThrGlyThrPtyrAsnGlnLeuGly 50
  |||||
150 TCTCGGCGCGGAGCGCGCATACCGGCACTGGTACACCAAGCTCGGC 199
  |||||
51 SerThrPheIleValIleThrAlaGlyAlaAspGlyAlaLeuThrGlyThr 67
  |||||
200 TCGACCTTCATCGTGAACGGGGCGCGGAGCGGCTGACCGGAACCTA 249
  |||||
67 rGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValLeuThrGlyArgT 84
  |||||
250 CGAGTCGGCGCTCGCAACGCGGAGCGGCTGCTGACCGGTCGTT 299
  |||||
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTTPThr 100
  |||||
300 ACACAGCGCCCGCGCGGCGGAGCGGAGCGGCGCCCTCGGTTGGAGC 349
  |||||
101 ValAlaTrpLysAsnAsnTyrrArgAsnAlaHisSerAlaThrThrTyrPse 117
  |||||
350 GTGGCGCTGGAAGATACCTACCGCAACGCCCACTCCGCGACCACTGGAG 399
  |||||
117 rGlyGlnTyrrValGlyAlaGlyAlaIleAsnThrGlnTrpLeuL 134
  |||||
400 CGGCGCATGAGTGGGGGGCGGCGGAGCGGAGATCAACACCCATGGCTGG 449
  |||||
134 eutHserGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
  |||||
450 TGCACCTCCGCGCACACCGGCGGAGCGGCTGGAAATCCAGCGCTGGTGGC 499
  |||||
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaI 167
  |||||
500 CAGGACACCTTCACCAAGGTGAAGCCGTCGCGCTCCATTCACACGGGG 549
  |||||
167 alySLysAlaGlyValAlaAsnAlaGlyAsnProLeuAspAlaValGlnGln 183
  |||||
550 GAAGAGAGCGCGGCTGACACAGCGCAACCGGCTGACGCGCTTACACAG 598
  |||||

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1986.DAT:N60626

seq\_documentation\_block:

ID N60626 standard; DNA: 1131 BP.

XX N60626;

DT 13-AUG-1991 (first entry)

DE Sequence of a portion of SA307 which encodes a streptavidin-like  
 DE polypeptide.

DE Antibiotic; biotin binding affinity; fusion protein; ds.

OS Streptomyces.

XX Key Location/Qualifiers

FT CDS 481..1029

FT /\*tag= a

FT /note= "P60623"

FT CDS 482..1030

FT /\*tag= b

FT /note= "P60624"

FT CDS 480..1031

FT /\*tag= c

FT /product= streptavidin-like polypeptide

FT /note= "P60625"

XX W08602077-A.

XX PD 10-APR-1986.

XX PF 01-OCT-1985; 85WO-0001901.

XX PR 02-OCT-1984; 84US-0656873.

XX PA (MEAD/) MEADE H M.

XX PI Meade HM, Garwin JL, Biogen NV;

XX DR WPI: 1986-106643/16.

XX P-PSDB; P60623, P60624, P60625.

XX PT DNA sequences and hybrid DNA sequences - encoding

XX PT streptavidin-like polypeptide, also joined to another protein,

XX PT e.g. tissue plasminogen activator

XX PS Disclosure; Fig. 2; 54pp; English.

XX CC The inventors claim the DNA sequence in SA307 which codes for a

XX CC streptavidin-like polypeptide (see N60626), and the polypeptide

XX CC encoded by it (P60625). They also claim hybrid SOs comprising N60626

XX CC and a second sequence coding for another protein, polypeptide,

XX CC peptide or AA (pref. tissue plasminogen activator (TPA)).

XX SO Sequence 1131 BP; 199 A; 409 C; 350 G; 173 T; 0 other:

## alignment\_scores:

Quality: 183.00 Length: 183  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-589-870-2 x N60626 ..

Align seg 1/1 to: N60626 from: 1 to: 1131

```

1 MetArgLysIleValValAlaIleAlaValSerLeuThrValSe 17
  |||||
480 ATGCCGACATGCTGTTGACCATCGCGTTCCCTGACCAACGGTCTC 529
  |||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGln 34
  |||||

```

```

|||||
530 GATTAGGGCAGGGCTTGGCAGACCCCTCCAAAGACTCGAAGCCGACG 579
34 aISeralAalaglualaglylierhGlyThrTrpTyraSngInleugly 50
|||||
580 TCTCGGCCCCCGAGGGCGGATCATCCGGCAGCGGTACCAACAGCTCGGCG 629
51 SerThPhleValThrAlaaglyAlaaspGlyAlaLeuThrGlyThrly 67
|||||
630 TCGACCTTATCTGTGACCGCGGGCGCGAGCGGGCCCTGATCGGAACCTA 679
67 rglUSeralAalaglYAsnAlaglUserArGlyrValleuThrGlyArG 84
|||||
680 CGAGTGGCGCGTGGCAACGCGAGAGCGGCTACGCTCGACGGGTCGT 729
84 yrasSerAlaProAlaThrAspGlySerGlyThraLeuglyTrpThr 100
|||||
730 ACGACAGCGCCCGCGCACCGACGCGAGCGGCGACCCCTCGGTTGAGCG 779
101 ValAlaTrpLysAsnAsnTyraTrpAsnAlaHisSerAlaThrThrTrpSe 117
|||||
780 GTGGCCTGGAGAGATTAATCAACGCAAGCCCACTCCGCAACACGTTGAG 829
117 rglYInTyraValGlyYAlaGlyAlaArGlyleAsnThrGlnTrpLeu 134
|||||
830 CGGCGGTACGTGCGGGCGCGGAGCGAGATCAACACCGAGTGGCTCG 879
134 euthSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
880 TGACCTCCGGCACACCGAGCCAGCCTGGAAGTCCACGCGTGTGCGCG 929
151 HisAspThrPherThrLysValLysProSerAlaAlaSerIleAspAla 167
|||||
930 CACGACACTTCAACAGGTGAAGCGGTCCGCGCTCCATCGACCGGCG 979
167 alySylalaglYAlaAsnAsnglyAsnProLeuAspAlaValGlnGln 183
|||||
980 GAAGAAGGCGGCGGTCAACACGCGCAACCCGCGTCCGATCACACAG 1028

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:291074
seq_documentation_block:
ID 291074 standard; DNA; 564 BP.
XX
AC 291074;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomycetes avidinii mst gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; mst gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomycetes avidinii.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
DR WPI: 2000-195402/17.
XX
P-PSDB: Y80513.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such

```

```

PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 85; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomycetes
CC avidinii mst gene as an example of a heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 564 BP; 109 A; 213 C; 169 G; 73 T; 0 other;

alignment_scores:
      quality: 182.00      length: 182
      ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 291074 ..

Align seg 1/1 to: 291074 from: 1 to: 564

2 ArgYlsIlaValValAlaAlaIleAlaValSerLeuThrValSerIle 18
|||||
14 CGAAGATGTCGTGACAGCATCGCCGTTCCCTGACCAACGCTCGAT 63
18 eThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValS 35
|||||
64 TAGGGCAGCGGCTTGGCAGACCCCTCCAAAGACTCGAAGGCCAGCTGT 113
35 eErAlaAlaaglualaglylierhGlyThrTrpTyraSngInleuglySer 51
|||||
114 CGGCGCGCGAGGGCGGATCATCCGGCAGCGGTGATCAACGCTCGGCTCG 163
52 ThrPhleIleValThrAlaaglyAlaaspGlyAlaLeuThrGlyThrTrp 68
|||||
164 ACCCTTATCTGTGACCGCGGGCGCGAGCGGCCCTGTACCGGAACCTAGA 213
68 uSerAlaValaglYAsnAlaglUserArGlyrValleuThrGlyArGTYra 85
|||||
214 GTGGCGCGTGGCAACGCGAGAGCGGCTACGTCCTGACCGGTCGTAGG 263
85 spSerAlaProAlaThrAspGlySerGlyThraLeuglyTrpThrVal 101
|||||
264 ACAGCGCCCGCGCACCGAGCGGCGACCGCCCTCGGTTGAGCGGTG 313
102 AlaTrpLysAsnAsnTyraTrpAsnAlaHisSerAlaThrThrTrpSerG 118
|||||
314 GCCTGGAAGATTAATCAACGCAAGCCCACTCCGCAACACGTTGAGCGG 363
118 yGlnTyraValGlyYAlaGlyAlaArGlyleAsnThrGlnTrpLeuTh 135
|||||
364 CCAGTACGTGCGGGCGCGGAGCGAGATCAACACCGAGTGGCTCGTA 413
135 hrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHis 151
|||||
414 CCTCGGCAACACCGAGCGGCTCGGAAGTCCACGCTGTGCGGCCAC 463
152 AspThrPherThrLysValLysProSerAlaAlaSerIleAspAlaIaly 168
|||||
464 GACACCTTCAACAGGTGAAGCGGTCCGCGCTCCATCGACGCGGCGA 513
168 slySylalaglYAlaAsnAsnglyAsnProLeuAspAlaValGlnGln 183
|||||

```

```

seq_name: /cgn2_2/gcdata/geneseq/geneseq/NA2000.DAT:291076
seq_documentation_block:
ID 291076 standard; DNA; 495 BP.
XX
XX 291076:
XX
XX 06-JUN-2000 (first entry)
XX
XX Streptomyces avidinii mprot gene.
DE
XX plant somatic tissue degeneration; plant essential factor; depletion;
XX viability; mprot gene; plant development; plant morphology; flower;
XX fruit plant; ds.
XX
XX Streptomyces avidinii.
OS
XX WO200007427-A2.
XX
XX 17-FEB-2000.
XX
XX 30-JUL-1999; 99WO-IL00420.
XX
XX 03-AUG-1998; 98IL-0125632.
XX
XX (AGRI-) AGRIC RES ORG.
XX
XX Kapulnik Y, Ginzberg I;
XX
XX WPI: 2000-195402/17.
XX
XX P-PSDB; Y80515.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
XX protein, useful for controlling plant development and morphology, such
XX as decreasing the number of flowers present to increase the number of
XX fruit -
XX
XX Examples: Page 86-87; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
XX plant tissue by expressing a heterologous protein capable of binding a
XX plant essential factor (PEF), in somatic plant tissue cells, where
XX heterologous protein expression causes depletion of the PEF so the plant
XX viability is maintained, while simultaneous degeneration of the somatic
XX plant tissue is effected. This sequence represents the Streptomyces
XX avidinii mprot gene as an example of a heterologous gene introduced
XX into the plants. The methods can provide for the selective and optionally
XX reversible cell degeneration in somatic plant tissue. They can be used
XX for artificially controlling plant development and morphology. They can
XX be used e.g. to decrease the number of fruits in fruit producing plants
XX so as to increase the number of fruits which reach maturity.
XX
XX Sequence 495 BP; 98 A; 188 C; 151 G; 58 T; 0 other;
SQ

alignment_scores:
Quality: 160.00 Length: 160
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 291076 ..

Align seg 1/1 to: 291076 from: 1 to: 495

24 AAtaSPProSeRtLySaSPSeRtLySaLaGInVaISeRtLaIaLaGInaLaG1 40
|||||
11 GCGAGACCCGTCGCAAGCACTCGAAGGCCCAAGTCTCGGCGCCGACGCGCG 60
|||||
40 ylleRrNglyThTTrPTyRaNgInLeuGlySeRThPhIleValThra 57
|||||

```

```

61 CATCACCGGCACTGTGTACAAACCAAGCTCGGCTGCAGCTTCACTGACGTGGACC 110
57 lAglyAlaAspGlyAlaLeuThrGlyThrTyGlnSerAlaValGlyAsn 73
111 CGGGGGCGGACGGCGCCCTGACCGGAACCTACGATCGGCGCTGGCAAC 160
74 lAAGUsErArGlyrValLeuThrGlyAlaGlyrYrAspSerAlaProAlaTh 90
161 GCCGGAGACCGCTACGTCCGACGGCGGTGTAGACAGGCCCGCCGGCAC 210
90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnT 107
211 CGACGGCAGCGGCGACCGCCCTCGCTTGGACGGTGGCTTGGAGATTAATCT 260
107 yrArGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGly 123
261 ACCGGAACGGCCACTCCGGGACCGACGAGGAGGGCGGACGTAAGTCGGGGC 310
124 lAAGlualAaTrgllEaSnThrGlnTrpLeuLeuThrSerGlyThrThG1 140
311 GCCGGGGGAGGATCAACACCCAGTGGCTGTGACTCCGGCACCAACCGA 360
140 uAlaSnAlaTrpLysSerThrLeuValGlyHisAspTrpPheThrLysV 157
361 GCCCAACGCGCTGGAAGTCCAGCGCTGGCGCCACGACACTTCCACCAAGG 410
157 alyrProSerAlaAlaSerlleAspAlaAlaLysLysAlaGlyValAlaSn 173
411 TGAAGCGCTCCGGCGGCTCATGACGGCGGCGGAGAAAGCGCGCTCAAC 460
174 AsnGlyAsnProLeuAspAlaValGlnGln 183
461 AACGGCAACCGGCTCGACGGCGGTTCAGCAG 490

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT.Q70659
seq_documentation_block:
ID Q70659 standard: DNA; 1356 BP.
XX
AC Q70659;
XX
DT 26-MAR-1995 (first entry)
XX
DE scfv PRAS108 and PRAS112 DNA.
XX
KW Amplification; single chain variable region fusion protein; PCR; ss
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 40..1348
XX FT /*tag= a
XX
NN W09415644-A.
XX
XX 21-JUL-1994.
XX
XX 17-JAN-1994; 94MO-GB00087.
XX
XX 15-JAN-1993; 93GB-0000686.
XX
XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
XX Deonarain M, Epenetos AA, Spooner RA;
XX
XX WPI; 1994-248907/30.
XX
XX P-PSDB; R56483.
XX
XX New cpgs, compiling a targeting portion and a cytotoxic portion
XX - used esp. for treating mammals for destroying target cells,
XX partic. tumour cells
XX
XX Claim 36; Fig 9; 114pp; English
XX

```

XX The sequence is that of the ScFv pRAS108 and pRAS112 between HindIII and  
 CC EcoRI site obtd. by PCR.  
 CC See also Q70652-69.  
 CC  
 XX

SQ Sequence 1356 BP; 332 A; 359 C; 359 G; 306 T; 0 other;

# alignment\_scores:

Quality:	160.00	Length:	160
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

# alignment\_block:

US-09-589-870-2 x Q70659 ..

Align seg 1/1 to: Q70659 from: 1 to: 1356

```

24 ALaasproSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaG1 40
|||||
865 GCAGACCGCTCCAGAGACTCCAAAGCTCAGGTTTCTGCAGCGGAAGCTGG 914
40 ylleThrGlyThrTrpTyraSngLleuGlySerThrPheIleValThrA 57
|||||
915 TATCACTGGACCTGGTATTAACCAACTGGGGCTGACTTTCATTGTGACCG 964
57 lAglyAlaAspGlyAlaLeuThrGlyThrTyrgLusSerAlaValGlyAsn 73
|||||
965 CTGGGCGGAGGAGCTGCTGACTGGACCTACGAAATCTGGGGTGGTAAAC 1014
74 AlagIusSerArgTyValLeuThrGlyValArgTyraSpsSerAlaProAlaTh 90
|||||
1015 GCAGAAATCCCGGTACTGACTGCGCGGTATATGACTCTGCAGCCGCCAC 1064
90 rAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTr 107
|||||
1065 CGATGGCTCTGGTACCGCTGGCTGGCTGACTGTGGTGAAGAAACAAC 1114
107 yTrArgSAlaHisSerAlaThrThrTrpSerGlyGlnTyraValGlyGly 123
|||||
1115 ATCGTAATGGCAGACGCGCTACGTGGCTGGCCAAATACCTGGCGGT 1164
124 AlagIuaAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG1 140
|||||
1165 GGTGAGGCTCGTATCAACACTCAGTGGCTGTTAATCATCGGAGCATACGGA 1214
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisSphrPheThrLysV 157
|||||
1215 AGCGAATGCATGAAATCGACACTAGTAGTGCATGACACCTTACCAAG 1264
157 allyProSerAlaHisSerIleAspAlaIalaLysAlaGlyValAsn 173
|||||
1265 TTAAACCTCTGCTGCTAGCATGTGATGCTGCCAAGAAAGACGAGCTTAAC 1314
174 AsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
1315 AACGGTAACCTCTAGACGCTTTTACAGCAA 1344

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA2000.DAT:Z91075

seq\_documentation\_block:

ID Z91075 standard; DNA; 492 BP.

XX AC Z91075;

DT 06-JUN-2000 (first entry)

DE Streptomyces avidinii prot gene.

XX Plant somatic tissue degeneration; plant essential factor; depletion;

KW viability; prot gene; plant development; plant morphology; flower;

KM fruit plant; ds.

XX

OS Streptomyces avidinii.

PN W0200007427-A2.

XX 17-FEB-2000.

PD 30-JUL-1999; 99W0-IL00420.

PR 03-AUG-1998; 98IL-0125632.

PA (AGRI-) AGRIC RES ORG.

PI Kapunlik Y, Ginzberg I;

DR WPI: 2000-195402/17.

DR P-PSDB; Y80514.

PT Degeneration of somatic plant tissue by expression of a heterologous protein, useful for controlling plant development and morphology, such as decreasing the number of flowers present to increase the number of fruit -

XX Examples; Page 86; 91pp; English.

CC The invention relates to a method of effecting degeneration of a somatic plant tissue by expressing a heterologous protein capable of binding a plant essential factor (PEF), in somatic plant tissue cells, where heterologous protein expression causes depletion of the PEF so the plant viability is maintained, while simultaneous degeneration of the somatic plant tissue is effected. This sequence represents the Streptomyces CC avidinii prot gene as an example of a heterologous gene introduced into the plants. The methods can provide for the selective and optionally CC reversible cell degeneration in somatic plant tissue. They can be used for artificially controlling plant development and morphology. They can be used e.g. to decrease the number of flowers in fruit producing plants CC so as to increase the number of fruits which reach maturity.

SQ Sequence 492 BP; 95 A; 188 C; 151 G; 58 T; 0 other;

# alignment\_scores:

Quality:	159.00	Length:	159
Ratio:	1.000 <td>Gaps:</td> <td>0 </td>	Gaps:	0
Percent Similarity:	100.000 <td>Percent Identity:</td> <td>100.000 </td>	Percent Identity:	100.000

# alignment\_block:

US-09-589-870-2 x Z91075 ..

Align seg 1/1 to: Z91075 from: 1 to: 492

```

25 AspproSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGly11 41
|||||
11 GACCCCTCCAGAGACTCGAAGGCCAGAGTCTGGCCCGCGAGGGCGGAT 60
41 eThrGlyThrTrpTyraSngLleuGlySerThrPheIleValThrAlaG 58
|||||
61 CACGCGACACTGCTGAACCACTCGGCTCGACCTTCATCGTGCAGCGGG 110
58 lYAlaAspGlyAlaLeuThrGlyThrTyrgLusSerAlaValGlyAsnAla 74
|||||
111 GCGCCGAGCGCGCCTGACCGGAACCTACGAGTCCGCGCGCAACGCC 160
75 GlusSerArgTyValLeuThrGlyArgTyraSpsSerAlaProAlaThrAs 91
|||||
161 GAGAGCCGCTACGTCTGACCGGTGTTACGACAGCGCCGCCACCGGA 210
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyra 108
|||||
211 CGGCAAGCGGACCGCCCTCGGTGGACGGTGGCTGGAAAGAAATACCTAC 260
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyraValGlyAla 124
|||||
261 GCAAGCCCACTCCGCGACACAGTGGAGGGCCAGTACGTGGCGGGCCG 310

```



```
125 GLUALAARGILEASNTHRGINTRPLEULEUTHRSERGIYTHRGLUAL 141
|||||
311 GAGCCAGAGATCAACACCCAGTGGCTGTGACCTCCGGCACCCAGGCG 360
141 AASNALATRPlySerThrLeuValGlyHisAspThrPheThrLysVal 158
361 CAACGGCTGGAAGTCCAGCTGTGTGGCCACGACACTTCACCAAGTGA 410
158 ySPROSERIAALASERTILEASPAALALYSLYSLAAGLYVALAASNA 174
411 AGCCGTCGCCCGCTTCATCGACGCGGAGAAAGCGCGCTCAACAAC 460
175 GLYASNPROLEUASPALAYALINGIN 183
461 GGCACCCGCTGCACGCCGTTCAGCAG 487

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1988.DAT: N80186

seq_documentation_block:
ID N80186 standard; DNA: 498 BP.
AC N80186;
XX
XX 13-OCT-1990 (first entry)
XX
XX Encodes biosynthetic protein with streptavidin trailer.
DE
XX Biosynthetic Antibody Binding site (BABS); variable domain;
KW anti-CPA monoclonal antibody; streptavidin; ss.
XX
XX Synthetic.
OS
XX W08809344-A.
XX
XX 01-DEC-1988.
XX
XX 19-MAY-1988; 88WO-US01737.
XX
XX 21-MAY-1987; 87US-0052800.
XX
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX
XX Huston JS, Oppermann H;
XX
XX WPI: 1988-353928/49.
XX
XX P-PSDB: P80160.
XX
XX Recombinant multifunctional protein -
XX having an antibody binding site and a sequence for biological activity,
XX ion sequestering or binding to a solid support.
XX
XX Disclosure: ; 115pp; English.
XX
XX Encodes multi-functional biosynthetic protein comprising single
XX chain BABS and streptavidin protein trailer linked via a spacer
XX sequence.
XX See also N80171-N80192.
XX
XX Sequence 498 BP; 117 A; 137 C; 133 G; 111 T; 0 other;

Alignment_scores:
Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-589-870-2 x N80186 ..
Align seg 1/1 to: N80186 from: 1 to: 498
25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI 41
```

```
|||||
13 GACCCGTCCAGAGACTCCAAAGCTCAGGTTTGTGTCGCCAAGCTGTA 62
41 eTHRGLYTHRTPTYRASGInLeuGlySerThrPheIleValThrAla 58
63 CACTGACACTGTGTATACCACTGGGGTCGACTTTCATGTGTACCGCTG 112
58 lVALASpGlyAlaLeuThrGlyThrTyrgIuSerAlaValGlyAAsNA 74
113 GTGCGAGAGAGCTGTGACTGGCAGCCTACGAATCTGCGGTGGTAACGA 162
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
163 GAATCCCGGTAGCTACTGACTGGCCGTTATGACTCGACCTCCACCGA 212
91 pGlySerGlyThrAlaLeuGlyTyrThrValAlaIatRPlyAsnAsnTyr 108
213 TGCTCTGTGTACCGCTCTGGGCTGGACTGTGGCTGGAAAAACACTATC 262
108 rGASNALAHISserAlaThrThrTrpSerGlyGlnTyrValGlyValAla 124
263 GTAATCGGCACAGCGCACTACGTGGTCTGGCCAAATACGTTGGCGGTCT 312
125 GLUALAARGILEASNTHRGINTRPLEULEUTHRSERGIYTHRGLUAL 141
313 GAGGCTCGTATCAACACTCAGTGGCTGTATACATCCGGCACTCCAGAG 362
141 AASNALATRPlySerThrLeuValGlyHisAspThrPheThrLysVal 158
363 GAATGCATGGAATCGACTAGTGCATGACACTTACCAAGTTA 412
158 ySPROSERIAALASERTILEASPAALALYSLYSLAAGLYVALAASNA 174
413 AGCCTTCTGCTGCTAGCATGTGATGCGCAAGAAAGCAGCGTTAAACAAC 462
463 GGTAAACCTCTAGACGCTGTTCAGCAA 489

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT: T75491

seq_documentation_block:
ID T75491 standard; DNA: 498 BP.
XX
XX T75491;
XX
XX 27-FEB-1998 (first entry)
XX
XX DNA for streptavidin.
XX
XX S-layer; sbs-A; vaccine; adjuvant; carrier; hybridisation assay;
XX molecular spinning nozzle; molecular laser; streptavidin; ds.
XX
XX Streptomyces avidinII.
OS
XX DE19603649-A1.
XX
XX 07-AUG-1997.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX (LOBIT/) LOBITZ W.
XX (SLEY/) SLEYTR U.
XX
XX Kuen B, Lubitz W, Sleytr U;
XX
XX WPI: 1997-394558/37.
XX
XX Preparation of S-layer proteins by expressing sbs-A gene in Gram
XX negative bacterium - or new sbs-B gene in any host, also new
XX recombinant proteins containing heterologous inserts, e.g.
PT
```

PT epitope(s), useful as vaccines and adjuvants

XX  
XX Example 8; Page 26; 31pp; German.

XX  
CC S-layer structures comprising the novel *Bacillus stearothermophilus*  
CC PV72 S-layer protein, sbs-A, can be used as vaccines or adjuvants,  
CC particularly when they include a bacterial ghost that may contain  
CC additional epitopes in its membrane. Other uses of recombinant  
CC sbs-A, depending on the nature of the inserted peptide, are as an  
CC universal carrier for biotinylated reactants for use in  
CC immunological or hybridisation assays (the insert is streptavidin,  
CC preferably encoded by the present sequence), to induce immune  
CC responses (epitopes), as a reagent for removing cytokine or toxin  
CC from serum (antigenic epitopes), as a molecular spinning nozzle  
CC (polyhydroxybutyrate synthase) and as a molecular laser  
CC (luciferase).

XX  
SO Sequence 498 BP; 120 A; 137 C; 131 G; 110 T; 0 other;

alignment\_scores:

Quality: 159.00 Length: 159  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x T75491 ..

Align seg 1/1 to: T75491 from: 1 to: 498

25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGlyI1 41  
|||||  
7 GACCCGTCACAGGACTCCAAAGCTCAGTTTCGACGCCAGAGCTGTAT 56  
41 ethrGlyThrTrrPtyrAsnGlnLeuGlySerThrPheIleValThlaG 58  
|||||  
57 CACTGGCACCTGGTATMACCACTGGGCTGACTGCTTCATGTGACCGCTG 106  
58 lValAspGlyAlaLeuThrGlyThrTyrGlySerAlaValGlyAla 74  
|||||  
107 GCGCGGACGAGCTCTGACTGGCACCCTACGAATCTCGGTGGTAAACGA 156  
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91  
|||||  
157 GAATCCCGCTAGCTACTGACTGGCGCTGTATGACTCGACCTGCCACGA 206  
91 polYSerGlyThrAlaLeuGlyTrrPthValAlaTrpLysAsnAsnTyrA 108  
|||||  
207 TGGCTCTGGTACCGCTCTGGCTGGACTGGCTGGGAAAAACAACCTATC 256  
108 rGAsnAlaHisSerAlaThrTrrTTPSerGlyGlnTyrValGlyAla 124  
|||||  
257 GTAAATGCGACAGCGCCACTACGTGCTGGCCAAATACCTTGGCGGTCT 306  
125 GluAlaArgIleAsnThrGlnTrrPleuLeuThrSerGlyThrThrGlnAl 141  
|||||  
307 GAGGCTCGTATCAACACACACGCGGTGTTAACATCGGCACATACGAAC 356  
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrIysValL 158  
|||||  
357 GAATGATGGAATGACACTAGTACGTATGACACCTTACCAAGTTA 406  
158 ySProSerAlaIleAsnThrIleAspAlaAlaLysLysAlaGlyValaAsnAs 174  
|||||  
407 AACCCTTCTCTCTAGCATTTGATGCTGCCAAGAAAGCAGCGCTAAACAC 456  
175 GlYAsnProLeuAspAlaValGlnGln 183  
|||||  
457 GGTAACCCCTCTAGACGCTGTCAGCAA 483

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1989.DAT:N90755

seq\_documentation\_block:

ID N90755 standard; DNA; 507 BP.  
XX  
AC N90755;  
XX  
DT 04-JUN-1990 (first entry)  
XX  
DE Sequence of streptavidin synthetic gene.  
XX  
KW Streptavidin; synthetic gene; Streptomyces avidinii;  
XX biotin; chimeric gene  
XX  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 10..492  
FT /tag= a  
FT 2..46  
FT misc\_feature  
FT /tag= b  
FT /label=BB214  
FT complement (6..53)  
FT /tag= c  
FT /label=BB215  
FT 47..86  
FT misc\_feature  
FT /tag= d  
FT /label=BB216  
FT complement (54..93)  
FT /tag= e  
FT /label=BB217  
FT 87..127  
FT misc\_feature  
FT /tag= f  
FT /label=BB218  
FT complement (94..134)  
FT /tag= g  
FT /label=BB219  
FT 128..171  
FT misc\_feature  
FT /tag= h  
FT /label=BB220  
FT complement (135..178)  
FT /tag= i  
FT /label=BB221  
FT 172..213  
FT misc\_feature  
FT /tag= j  
FT /label=BB222  
FT complement (179..220)  
FT /tag= k  
FT /label=BB223  
FT 214..256  
FT misc\_feature  
FT /tag= l  
FT /label=BB224  
FT complement (221..268)  
FT /tag= m  
FT /label=BB225  
FT 257..295  
FT misc\_feature  
FT /tag= n  
FT /label=BB226  
FT complement (264..302)  
FT /tag= o  
FT /label=BB227  
FT 296..338  
FT misc\_feature  
FT /tag= p  
FT /label=BB228  
FT complement (303..345)  
FT /tag= q  
FT /label=BB229  
FT 399..380  
FT misc\_feature  
FT /tag= r  
FT /label=BB230  
FT complement (346..387)  
FT /tag= s  
FT /label=BB231  
FT 381..423  
FT misc\_feature  
FT /tag= t  
FT /label=BB232  
FT complement (390..431)  
FT /tag= u

```

163  GAAATCCGCGCTACGTACTGACGTGGCCGTTATGACTCTGCACCTGCCACCGCA 212
91  polYserGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTrpAla 108
|||||
213  TGGCTCTGTACCGCTCTGGGGCTGGAAGTGGCTTGGAAAAACAACATTC 262
108  rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTrpValGlyValAla 124
|||||
263  GTAATGCCACAGCGCCACTACGTGCTGGCCCAATACCTTGGCGTCT 312
125  GUAAlaArgIleAsnTrpGlnTrpLeuLeuThrSerGlyThrThrGlnAl 141
|||||
313  GAGGCTCGTATCAACACTCAAGTGCTGTTAACATCCGGACATACCGAAGC 362
141  aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
363  GAATGCATGGAATCGACACTAGTAGGTCTATGACACCTTATACCAAGTTA 412
158  ySProsSerAlaAlaSerIleAspAlaAlaLysIleAspGlyValAsnAs 174
|||||
413  AGCCCTTCGCTGCTAGCATTTGATGCTGCCAAGAAACAGACGCTAAACAAC 462
175  GlYAsnProLeuAspAlaValGlnGln 183
463  GGTAAACCTCTTAGACGCTGTTCAGCA 489

seq_name= /cgn2_2/gcdata/geneseq/geneseqn/NA1993.DAT:053412

seq_documentation_block:
ID  Q053412 standard; DNA; 552 BP.
XX
AC  Q053412;
XX
DT  27-JUN-1994 (first entry)
XX
XX  Streptavidin gene.
XX
XX  Streptavidin; protein secretion; Bacillus subtilis; ss.
XX
XX  Streptomyces avidinII.
XX
OS  Key Location/Qualifiers
FH  CDS 1..552
FT  mat_peptide 73..519
FT  sig_peptide 1..72
FT  misc_feature 109..519
FT  /tag= d
FT  /note= "fused to npr signal peptide in pBB673"
XX
XX  W09324631-A.
XX
XX  09-DEC-1993.
XX
XX  27-MAY-1993; 93MO-US05240.
XX
XX  29-MAY-1992; 92US-0891524.
XX
XX  (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX  Nagarañan V.
XX
XX  WPI; 1993-405822/50.
XX
XX  P-PSDB; R44491.
XX
XX  Streptavidin prodn. from Bacillus subtilis - using signal protein
XX  from bacterial exo-protein and expression element from Gram
XX  positive bacterial protein.
XX
XX  Disclosure; Fig 1b; 54pp; English.
XX

```

CC Tetrameric biologically active streptavidin is produced by secretion  
CC from *Bacillus subtilis* transformed with a plasmid encoding the  
CC sequence.

**SQ** Sequence 552 BP; 105 A; 207 C; 167 G; 70 T; 3 other;

```
alignment_scores:      Length: 157
      Quality:
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.0000
```

```
alignment_block:
US-09-589-870-2 x Q53412 ..
```

Align seg 1/1 to: Q53412 from: 1 to: 552

27 SerLyAspSerLysAlaGlnValSerAlaAlaGlnAlaGlyIleThrG1 43  
79 TCCAAAGACTCCAAAGGCCAAGGTCCTGGCGCGCCGAGCGCCGATCCGG 128  
43 yHrTrpTyraSngInLeuGlySerThrPheIleValThrAlaGlyAla 60  
129 CACCTGGTACCAACAGACTCGCTGCACCTTATCTGTACCGCGGGCCGG 178  
60 spGIyAlaLeuThrGlyThrTyrgInuSerAlaValAlaGlyAsnAlaGluSer 76  
179 ACGGCCCGCTGCACGGAACTTCAGAGTCGCGCGTGGCAACGCCGAGAGC 228  
77 ArgTyValValLeuThrGlyArgTyrsPserAlaProAlaThrAspGlySe 93  
229 CGGTACGGCTTCAACGGTCGTGTACACAGCGCCCGCGCACCGAGCGAG 278  
93 rGIyThrAlaLeuGlyTrpThValAlaTrpLysAsnAntTyraArgAsnA 110  
279 CGGCACCGCCCTCGGTGGACCGTGGCTGGAAAGATATACCTACCGCAACG 328  
110 IaHisSerAlaThrThrTrpSerGlyGlnTyValGlyAlaGlyAlaGluAla 126  
329 CCCACTCCCGCGACCAAGTGAGCGGCCCATGATCGTGGCGCGCGGAGGCG 378  
127 ArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAlaAsnAl 143  
379 AGGATTAACAACCCAGAGGGCTGTGACTTCGCGGACACACCGAGGCCAACGC 428  
143 aTrpLysSerThrIleValGlyHisAspThrPheThrLysValLysProS 160  
429 CTGGAAGTCCACGCTGGTGGCGCCACGAACTTCAACCAAGGTGAAGCGCT 478  
160 eAlaIaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsn 176  
479 CCGCGCGCTTCATCGACGCGGGAAGAAAGCGCGGTCAACACAGGCAAC 528  
177 ProLeuAspAlaValGlnGln 183  
529 CCGCTTGACGCGGTTCAGAGC 549

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q706660

```
seq_documentation_block:
ID    Q70660 standard; DNA; 1296 BP
```

AC Q70660;

DT 26-MAR-1995 (first entry)

DE SCFV PRAS109 and PRAS113 DNA.

**KW** Amplification; single chain variable region fusion protein; PCR; ss.

OS Synthetic.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	CDS	40..1290
FT		/*tag= a

PN W09415644-A.

PD 21-JUL-1994.

PF 17-JAN-1994; 94WO-GB00087.

PR 15-JAN-1993; 93GB-0000686.

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

PI Deonarain M, Epenetos AA, Spooner RA;

DR WPI; 1994-248907/30.

XX

PT - used esp. for treating mammals for destroying target cells,

XX  
XX

XX

CC EcORI site obtd. by PCR.

XX

alignment\_scores:

```
alignment_scores:
  Quality: 140.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Length: 140
  Gaps: 0
  Percent Identity: 100.0000
```

alignment\_block:  
US-09-589-870-2 x Q70660

Align seg 1/1 to: Q70660 from: 1 to: 1296

[illegible]

157 allYpRoSeRAlaAlSeR 163  
|||||  
1265 TTAGCCTTCTGCTGCTAGC 1284

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 01:07:06 ; Search time 63.76 Seconds  
(without alignments)  
194.884 Million cell updates/sec

Title: US-09-589-870-2

Perfect score: 183

Sequence: 1 MKRIVVAALVSLTTSITA.....IDAKKAGVNNGNPLDAVQ 183

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 195891 seqs, 67900655 residues

Word size : 0

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: PIR66:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	183	2	A23513 streptavidin precu
2	123	67.2	183	2	S57284 streptavidin v1 pr
3	59	32.2	183	2	S57285 streptavidin v2 pr
4	8	4.4	212	2	T46600 nitrogen metabolic
5	8	4.4	366	2	F81578 DNA polymerase III
6	8	4.4	366	2	C81713 DNA polymerase III
7	8	4.4	366	2	H72090 DNA polymerase III
8	8	4.4	367	2	B83186 probable RND efflu
9	8	4.4	416	2	E71559 probable DNA pol I
10	8	4.4	470	2	D75375 conserved hypochet
11	8	4.4	488	1	S11910 nitrogen metabolic
12	8	4.4	1067	2	G81825 probable drug effl
13	8	4.4	1067	2	E81051 multiple transfera
14	8	4.4	1067	2	E81051 DNA topoisomerase
15	8	4.4	1067	2	S59969 lysin protein S -
16	8	4.4	107	1	YVBP1 lysin protein S -
17	8	3.8	108	1	YVBP2 ribosomal protein
18	8	3.8	133	2	S29885 ribosomal protein
19	8	3.8	149	2	C71632 hypothetical prote
20	8	3.8	149	2	F70599 hypothetical prote
21	8	3.8	152	2	S36108 superoxide dismuta
22	8	3.8	152	2	S72235 superoxide dismuta
23	8	3.8	164	2	T35813 hypothetical prote
24	8	3.8	166	2	B72607 hypothetical prote
25	8	3.8	182	2	E82235 hypothetical prote
26	8	3.8	191	2	T44563 probable thiamin b
27	8	3.8	218	2	S51751 dihydrodipicolinat
28	8	3.8	225	2	T33055 hypothetical prote
29	8	3.8	255	1	CPSMWU muramoylpentapepti

30	7	3.8	259	2	D83216 probable aldolase
31	7	3.8	266	1	ELRT1 pancreatic elastas
32	7	3.8	275	2	B4394 translation elonga
33	7	3.8	295	2	T00399 transcription fact
34	7	3.8	300	2	A55792 beta-lactamase (EC
35	7	3.8	319	2	A70605 probable acyl-coa
36	7	3.8	345	2	T08563 dnaJ-related prote
37	7	3.8	376	2	JC6535 multiubiquitin-pro
38	7	3.8	377	2	S63671 26S proteinase cha
39	7	3.8	378	2	S71201 biotin synthase (E
40	7	3.8	397	2	T35609 white protein I - S
41	7	3.8	413	2	F69007 translation elonga
42	7	3.8	433	2	S55968 probable membrane
43	7	3.8	437	2	C72570 probable translati
44	7	3.8	450	1	S13730 pma protein - Esc
45	7	3.8	453	2	T38707 probable initiator

#### ALIGNMENTS

RESULT 1  
A23513 streptavidin precursor - Streptomyces avidinii  
C:Species: Streptomyces avidinii  
C:Date: 03-Nov-1987 #sequence=1 #revision 03-Nov-1987 #text\_change 10-Dec-1999  
C:Accession: A23513; S11540  
R:Argarana, C.E.; Kuntz, I.D.; Birken, S.; Axel, R.; Cantor, C.R.  
Nucleic Acids Res. 14, 1871-1882, 1986  
A:Title: Molecular cloning and nucleotide sequence of the streptavidin gene.  
A:Reference number: A23513; MUID:86148514  
A:Accession: A23513  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-183 <ARG>  
A:Cross-references: GB:X03591; NID:946740; PIDN:CAA27265.1; PID:946741  
R:Gittlin, G.; Bayer, E.A.; Mallick, M.  
Biochem. J. 269, 527-530, 1990  
A:Title: Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine res  
A:Reference number: S11540; MUID:90351377  
A:Accession: S11540  
A:Molecule type: protein  
A:Residues: 54-66, 'X', 68 <GIT>  
C:Superfamily: streptavidin  
C:Keywords: biotin; homotetramer  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-183/Product: streptavidin #status predicted <MAT>  
F:67/Binding site: biotin (Tyr) #status predicted  
F:78/Binding site: biotin (Tyr) #status predicted

Query Match 100.0%; Score 183; DB 2; Length 183;  
Best local Similarity 100.0%; Pred. No. 7.4e-172;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRKIIVAAIVSLTTSITASADPSKDSKAOVSAEAGITGTWYNQSGFTVITAGAD	60
DB	1	MRKIIVAAIVSLTTSITASADPSKDSKAOVSAEAGITGTWYNQSGFTVITAGAD	60
QY	61	GALTGTYESAVNAESRYVLTGRYDSAPATDGSGLTALGWTAVANKNNYRNAHSATTSQGY	120
DB	61	GALTGTYESAVNAESRYVLTGRYDSAPATDGSGLTALGWTAVANKNNYRNAHSATTSQGY	120
QY	121	VGAEARINTQWLLTSGTTTANAMKSTLVGHDPFTYKPSAASIDAARKAGVNNGNPLDA	180
DB	121	VGAEARINTQWLLTSGTTTANAMKSTLVGHDPFTYKPSAASIDAARKAGVNNGNPLDA	180
QY	181	VQO 183	
DB	181	VQO 183	

RESULT 2

S57284  
streptavidin v1 precursor - Streptomyces venezuelae  
C:Species: Streptomyces venezuelae  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 10-Dec-1999  
C:Accession: S57284  
R:Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.  
Biochim. Biophys. Acta 1263, 60-66, 1995  
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom  
A:Reference number: S57284; MUID:95359204  
A:Accession: S57284  
A:Molecule type: DNA  
A:Residues: 1-183 <BAV>  
A:Experimental source: strain Tue2460  
C:Superfamily: streptavidin  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-183/Product: streptavidin v1 #status predicted <MAT>

Query Match	67.2%	Score 123	DB 2	Length 183
Best Local Similarity	100.0%	Pred. No. 5.3e-113		
Matches 123	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	MRIYVAIAIVSLTFTVSTTASASADPSKDSKROVBAEAGITGTVYNOLGSEFTIYAGAD	60		
-				
Db 1	MRIYVAIAIVSLTFTVSTTASASADPSKDSKROVBAEAGITGTVYNOLGSEFTIYAGAD	60		
QY 61	GALGTGYESAAGNRESRYLTGRYDSAPATDGSCTALGTVAMKNNYRNAHSATTWSGQY	120		
Db 61	GALGTGYESAAGNRESRYLTGRYDSAPATDGSCTALGTVAMKNNYRNAHSATTWSGQY	120		
QY 121	VGG 123			
Db 121	VGG 123			

S57285 streptavidin v2 precursor - Streptomyces venezuelae  
C/Species: Streptomyces venezuelae  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 10-Dec-1999  
C/Accession: S57285  
R/Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.  
Biochim. Biophys. Acta 1263, 60-66, 1995  
A>Title: Close similarity among streptavidin-like, biotin-binding proteins from streptomycetes  
A/Reference number: S57284; MUID:95359204  
A/Accession: S57285  
A/Molecule type: DNA  
A/Residues: 1183 <BAY>  
A/Experimental source: strain Tue2605  
A/Superfamily: streptavidin  
F/1-24/Domain: signal sequence #status predicted <SIG>  
F/25-183/Product: streptavidin v2 #status predicted <MAT>

	Query Match	32.2%	Score 59	DB 2	Length 183
Best Local Similarity	100.0%	Pred.	No. 3.2e-50		
Matches 59	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
63	LGCTYSAVGNASRIVLTGRDSPAATDGGTALGWTAVMKNNYRNAHSATTWGGQYV				
63	LGCTYSSAAGNASESRVLTGRDSPAATDGGTALGWTAVMKNNYRNAHSATTWGGQYV				

RESULT 4  
T46600  
nitrogen metabolic regulation protein nmr [imported] - Neurospora crassa  
C/Species: Neurospora crassa  
C/Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 02-Sep-2000  
C/Accession: T46600  
R/Tarral, G.; Marziuf, G.A.  
Curr. Genet. 20, 283-286, 1991  
A/Tittle, generation of new mutants of nmr, the negative-acting nitrogen regulatory gene  
#Reference number: Z23098; MUID:92035098

A:Accession: T46600  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-212 <JAN>  
 A:Cross-references: EMBL:S64286; PIDN:CAB32318.1  
 C:genetics:  
 A:Gene: nm1  
 C:Superfamily: nitrogen metabolic regulation protein nm1r

Query March	4.4%	Score 8:	DB 2:	Length 212;
Best Local Similarity	100.0%	Pred. No. 3.8;		
Matches	8;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
QY	165	DAAKRAGV	172	
Db	83	DAAKRAGV	90	

RESULT 5  
F81578  
DNA polymerase III, beta chain CP0419 [imported] - Chlamydophil pneumoniae (strain A  
C:Species: Chlamydophil pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Sep-2000  
C:Accession: F81578  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: F81578  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <REA>  
A:Cross-references: GB:AE002203; GB:AE002161; NID:g7198339; PIDN:AAF38262.1; PID:g718  
C:Genetics:  
A:Experimental source: strain AR39, HL cells  
A:Gene: CP0419  
A:Superfamily: DNA-directed DNA polymerase III beta chain

	Query Match	4.43; Score 8; DB 2; Length 366;
	Similarity	100.0%;
	Best Local	Pred. No. 6;
	Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	75 ESRVLTG 82	
db	149 ESRVLTG 156	

RESULT 6  
 C81713  
 DNA polymerase III, beta chain TC0347 [imported] - Chlamydia muridarum (strain Nigg)  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Sep-2000  
 C:Accession: C81713  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Debony, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
 A:Reference number: AB1500; MUID:20150255  
 A:Accession: C81713  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-366 <TET>  
 A:Cross-references: GB:AE002302; GB:AE002160; NID:g7190382; PIDN:AAF39208.1; PID:g719  
 A:Experimental source: strain Nigg (Mopn)  
 C:Genetics:  
 A:Gene: TC0347  
 A:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.48; Score 8; DB 2; Length 366;



```

Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 ESRVLTG 82
|||||
Db 149 ESRVLTG 156

RESULT 7
H72090
DNA polymerase III (beta chain) - Chlamydomonas reinhardtii (strain CWI029)
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: H72090
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: H72090
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-366 <ARN>
A:Cross-references: GB:AE001618; GB:AE001363; NID:94376610; PIDN:AA18487.1; PID:9437661
A:Experimental source: strain CWI029
C:Genetics:
A:Gene: dan
C:superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.4%; Score 8; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 ESRVLTG 82
|||||
Db 149 ESRVLTG 156

RESULT 8
B83186
probable RND efflux membrane fusion protein precursor PA3677 [Imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: B83186
R:Storer, C.K.; Pham, X.Q.; Ewvin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950
A:Accession: B83186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-367 <STO>
A:Cross-references: GB:AE004787; GB:AE004091; NID:99949835; PIDN:AG07065.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3677

Query Match 4.4%; Score 8; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 AOVSAEA 39
|||||
Db 103 AOVSAEA 110

RESULT 9
E71559
probable DNA pol III (beta chain) - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis

```

```

C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 11-Jan-2000
C:Accession: E71559
R:Stephens, R.S.; Kallman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: E71559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <ARN>
A:Cross-references: GB:AE001282; GB:AE001273; NID:93328466; PIDN:AA67666.1; PID:9332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: dan
C:superfamily: DNA-directed DNA polymerase III beta chain

Query Match 4.4%; Score 8; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 ESRVLTG 82
|||||
Db 199 ESRVLTG 206

RESULT 10
D75375
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: D75375
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
., M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: D75375
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-470 <WHI>
A:Cross-references: GB:AE002004; GB:AE005113; NID:96459366; PIDN:AA11162.1; PID:9645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1599
A:Map position: 1

Query Match 4.4%; Score 8; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AGADGALT 64
|||||
Db 392 AGADGALT 399

RESULT 11
S11910
nitrogen metabolic regulation protein nmr - Neurospora crassa
N:Alternate names: nmr protein
C:Species: Neurospora crassa
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S11910; S11924
R:Young, J.L.; Jarai, G.; Fu, Y.H.; Marzluf, G.A.
Mol. Gen. Genet. 222, 120-128, 1990
A:Title: Nucleotide sequence and analysis of NMR, a negative-acting regulatory gene 1
A:Reference number: S11910; MUID:91042412
A:Accession: S11910
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-488 <YOU>

```

C:Genetics:  
A:Gene: nmr  
C:Superfamily: nitrogen metabolic regulation protein nmr  
C:Keywords: nucleus; transcription regulation

Query Match 4.4%; Score 8; DB 1; Length 488;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 DAAKAGV 172  
|||||||  
DB 181 DAAKAGV 188

RESULT 12  
G81825  
probable drug efflux protein NMA1969 [imported] - Neisseria meningitidis (group A strain

C:Species: Neisseria meningitidis  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 19-May-2000

C:Accession: G81825

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

R:Holroyd, S.; Jagals, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:20222556

A:Accession: G81825

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1067 <PAR>

A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85189.1; PID:g738060

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMA1969

C:Superfamily: acriflavin resistance protein

Query Match 4.4%; Score 8; DB 2; Length 1067;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 AAEAGITG 43  
|||||||  
DB 1054 AAEAGITG 1061

RESULT 13

E81051  
multiple transferable resistance system protein MtrD NMB1715 [imported] - Neisseria men

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000

C:Accession: E81051

R:Retellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Masiarant, V.; Pizsa, M.;

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: E81051

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1067 <TEP>

A:Cross-references: GB:AE002521; GB:AE002098; NID:g7226962; PIDN:AAFA2062.1; PID:g722697

A:Experimental source: serogroup B, strain MKD58

C:Genetics:

A:Gene: NMB1715

C:Superfamily: acriflavin resistance protein

Query Match 4.4%; Score 8; DB 2; Length 1067;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 AAEAGITG 43  
|||||||  
DB 1054 AAEAGITG 1061

RESULT 14

S59969  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) isoform beta - Chinese hamster

M:Alternate names: DNA topoisomerase II isoform beta; DNA gyrase

C:Species: Crictetus griseus (Chinese hamster)

C>Date: 15-Feb-1996 #sequence\_revision 13-Mar-1997 #text\_change 16-Jul-1999

C:Accession: S59969; S54154

R:Derudder, S.; Frey, S.; Delaporte, C.; Jacquemin-Sablon, A.

Biochim. Biophys. Acta 1264, 178-182, 1995

A:Title: Cloning and characterization of full-length cDNAs coding for the DNA topois

A:Reference number: S59969; MUID:96085121

A:Accession: S59969

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-1612 <DER>

A:Cross-references: EMBL:X86455; NID:g790987; PIDN:CAA60173.1; PID:g790988

A:Experimental source: Lung

C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-

C:Keywords: ATP; DNA binding; DNA replication; isomerase; nucleus

F:697-927/Domain: phage T4 DNA topoisomerase (ATP-hydrolyzing) medium chain homology

Query Match 4.4%; Score 8; DB 2; Length 1612;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AGADGALT 64  
|||||||  
DB 7 AGADGALT 14

RESULT 15

YBPL  
lysis protein S - phage lambda

C:Species: phage lambda

C>Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 23-Jul-1999

C:Accession: H94164; G43012; A04327

R:Daniels, D.

submitted to the Nucleic Acid Sequence Database, September 1982

A:Reference number: A94614

A:Accession: H94164

A:Molecule type: DNA

A:Residues: 1-107 <DNA>

A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g21510

A:Title: Nucleotide sequence of bacteriophage lambda DNA.

A:Reference number: A92891; MUID:83189071

A:Accession: G43012

A:Molecule type: DNA

A:Residues: 1-107 <SAN>

A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g21510

A>Note: the lambda DNA sequence is from the standard strain lambda-clint857sam7; t

C:Genetics:

A:Gene: S

A:Map position: 93.16-93.82

C:Superfamily: phage lambda lysis protein S

C:Keywords: host cell lysis

Query Match 3.8%; Score 7; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAKKAGV 172  
|||||||  
DB 95 AAKKAGV 101

Fri Apr 6 08:41:00 2001

us-09-589-870-2.rpr

Page 5

Search completed: April 6, 2001, 01:26:46  
Job time: 1180 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 01:22:46 ; Search time 67.41 Seconds

(Without alignments)  
87.670 Million cell updates/sec

Title: US-09-589-870-2

Sequence: 1 MRKIYVAIAIVSLTIVSITA.....IDAKKAGVNNNGNPLDAVOQ 183

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 88757 seqs, 32294092 residues

Word size : 0

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	183	1	SAV_STRAV
2	123	67.2	183	1	SAV_STRVL
3	59	32.2	183	1	SAV2_STRVL
4	8	4.4	488	1	NMR_NEUCR
5	8	4.4	1612	1	TP28_CRITLO
6	7	3.8	107	1	VLVS_LAMB
7	7	3.8	108	1	VLVL_BPP22
8	7	3.8	132	1	RS8_MICLU
9	7	3.8	151	1	SOD4_MAIZE
10	7	3.8	151	1	SOD5_MAIZE
11	7	3.8	152	1	SODC_CAVPO
12	7	3.8	255	1	CBPM_STRAL
13	7	3.8	266	1	ELI_RAT
14	7	3.8	302	1	DAPA_PROMA
15	7	3.8	376	1	PSD4_MOUSE
16	7	3.8	377	1	PSD4_MOUSE
17	7	3.8	378	1	BIOB_HUMAN
18	7	3.8	379	1	RTC2_CABEL
19	7	3.8	397	1	WH42_STRCO
20	7	3.8	413	1	EFIA_METH
21	7	3.8	437	1	EFIA_MERPE
22	7	3.8	450	1	PMB4_ECOLI
23	7	3.8	453	1	YAW6_SCHPO
24	7	3.8	474	1	GSAL_ARATH
25	7	3.8	553	1	DPOL_BPPRD
26	7	3.8	583	1	ARA2_ECOLI
27	7	3.8	637	1	NU5M_STRPU
28	7	3.8	681	1	TKT2_YEAST
29	7	3.8	737	1	CATA_YERPE
30	7	3.8	783	1	OPT3_CABEL
31	7	3.8	960	1	FSD1_MOUSE
32	7	3.8	1025	1	MK21_YEAST
33	7	3.8	3010	1	POLG_HCVBK

#### ALIGNMENTS

RESULT ID	SAV_STRAV	STANDARD	PRT	183 AA.
AC	P22629;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	STREPTAVIDIN PRECURSOR.			
OS	Streptomyces avidinii.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomyces; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 24-64.			
RX	MEDLINE=86148514; PubMed=3951999;			
RA	Agarana C.E., Kuntz I.D., Birken S., Axel R., Cantor C.R.;			
RT	"Molecular cloning and nucleotide sequence of the streptavidin gene.";			
RL	Nucleic Acids Res. 14:1871-1882(1986).			
RN	[2]			
RP	IMPORTANCE OF TRP IN BIOTIN-BINDING.			
RX	MEDLINE=89134083; PubMed=3223904;			
RA	Gittlin G., Bayer E.A., Wilchek M.;			
RT	"Studies on the biotin-binding site of streptavidin. Tryptophan residues involved in the active site.";			
RL	Biochem. J. 256:279-282(1988).			
RN	[3]			
RP	IMPORTANCE OF TYR IN BIOTIN-BINDING.			
RX	MEDLINE=90351377; PubMed=2386489;			
RA	Gittlin G., Bayer E.A., Wilchek M.;			
RT	"Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine residues are involved in the binding site.";			
RL	Biochem. J. 269:527-530(1990).			
RN	[4]			
RP	CELL-BINDING.			
RX	MEDLINE=90358825; PubMed=2390089;			
RA	Alon R., Bayer E.A., Wilchek M.;			
RT	"Streptavidin contains an RVD sequence which mimics the RGD receptor domain of fibronectin.";			
RL	Biochem. Biophys. Res. Commun. 170:1236-1241(1990).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY OF 37-157.			
RX	MEDLINE=89100243; PubMed=2911722;			
RA	Weber P.C., Ohlendorf D.H., Wendoloski J.J., Salemme F.R.;			
RT	"Structural origins of high-affinity biotin binding to streptavidin.";			
RL	Science 243:85-88(1989).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.			
RX	MEDLINE=97337436; PubMed=9194176;			
RA	Freitag S., le Trong I., Klumb L., Stayton P.S., Stenkamp R.E.;			
RT	"Structural studies of the streptavidin binding loop.";			
RL	protein Sci. 6:1157-1166(1997).			
RN	[7]			
RP	X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS) OF 37-157.			
RX	MEDLINE=97294734; PubMed=9148939;			
RA	Katz B.A., Cass R.T.;			
RT	"In crystals of complexes of streptavidin with peptide ligands containing the HPQ sequence the pKa of the peptide histidine is less			

34	7	3.8	3011	1	POLG_HCV1	P26664 h genome po
35	6	3.3	17	1	UP36_UPEMJ	P82043 uperolela m
36	6	3.3	61	1	PSA_ERMHE	O54457 erwina her
37	6	3.3	61	1	RCRO_BPP22	P09964 bacterioph
38	6	3.3	81	1	YMAA_ECOLI	P76036 escherichia
39	6	3.3	115	1	VS_BPT4	P13310 bacterioph
40	6	3.3	117	1	NLTP_SPIOL	P10976 spinacia ol
41	6	3.3	119	1	R26A_YEAST	P39938 saccharomyc
42	6	3.3	120	1	SODC_ASPUA	O12548 aspergillus
43	6	3.3	121	1	VAIF_CABEL	O23680 caenorhabdl
44	6	3.3	121	1	Y20L_ARATH	O80813 arabidopsis
45	6	3.3	122	1	H2B1_TIGCA	P35068 tigrilopus c

RT than 3.0.";  
RL J. Biol. Chem. 272:13220-13228(1997).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 37-157.  
RX MEDLINE=98070615; PubMed=9405158;  
RA Katz B.A.;  
RT "Binding of biotin to streptavidin stabilizes intersubunit salt  
RT bridges between Asp61 and His87 at low pH.";  
RL J. Mol. Biol. 274:776-800(1997).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 37-157 OF MUTANTS.  
RX MEDLINE=98300340; PubMed=9636711;  
RA Freitag S., le Trong I., Chikoti A., Klumb L.A., Stayton P.S.,  
RA Stenkamp R.E.;  
RT "Structural studies of binding site typtophan mutants in the high-  
RT affinity streptavidin-biotin complex.";  
RL J. Mol. Biol. 279:211-221(1998).  
CC -I- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.  
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE  
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).  
CC -I- SUBUNIT: HOMOTETRAMER.  
CC -I- SUBCELLULAR LOCATION: SECRETED  
CC -I- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.  
CC -I- DATABASE: NAME=Prozyme technical fact sheet;  
CC WWW="http://www.prozyme.com/technical/sai0data.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, X03591; CAA27265.1; -  
DR PIR, A23513; A23513.  
DR PDB, 1STP; 15-OCT-94.  
DR PDB, 1PTS; 15-OCT-94.  
DR PDB, 1SRE; 01-NOV-94.  
DR PDB, 1SRF; 01-NOV-94.  
DR PDB, 1SRG; 01-NOV-94.  
DR PDB, 1SRH; 01-NOV-94.  
DR PDB, 1SRI; 01-NOV-94.  
DR PDB, 1SRJ; 01-NOV-94.  
DR PDB, 1RST; 15-MAY-97.  
DR PDB, 1RSU; 15-MAY-97.  
DR PDB, 1STD; 03-APR-96.  
DR PDB, 1SLE; 03-APR-96.  
DR PDB, 1SLF; 03-APR-96.  
DR PDB, 1SLG; 03-APR-96.  
DR PDB, 1STR; 08-MAR-96.  
DR PDB, 1STS; 08-MAR-96.  
DR PDB, 1SMA; 04-MAR-98.  
DR PDB, 1SWB; 04-MAR-98.  
DR PDB, 1SWC; 04-MAR-98.  
DR PDB, 1SWD; 04-MAR-98.  
DR PDB, 1SWE; 04-MAR-98.  
DR PDB, 1SMG; 15-JUL-98.  
DR PDB, 1SMH; 23-MAR-99.  
DR PDB, 1SMJ; 23-MAR-99.  
DR PDB, 1SMK; 23-MAR-99.  
DR PDB, 1SWL; 23-MAR-99.  
DR PDB, 1SWN; 23-MAR-99.  
DR PDB, 1SMO; 23-MAR-99.  
DR PDB, 1SWP; 23-MAR-99.  
DR PDB, 1SWQ; 23-MAR-99.  
DR PDB, 1SWR; 23-MAR-99.  
DR PDB, 1VMA; 18-MAR-98.  
DR PDB, 1VMB; 18-MAR-98.  
DR PDB, 1VMC; 18-MAR-98.  
DR PDB, 1VMD; 18-MAR-98.  
DR PDB, 1WE; 18-MAR-98.  
DR PDB, 1WVF; 18-MAR-98.

DR PDB, 1VMG; 18-MAR-98.  
DR PDB, 1VMH; 18-MAR-98.  
DR PDB, 1VMI; 18-MAR-98.  
DR PDB, 1VMJ; 18-MAR-98.  
DR PDB, 1VMK; 18-MAR-98.  
DR PDB, 1VWL; 18-MAR-98.  
DR PDB, 1VMN; 18-MAR-98.  
DR PDB, 1VMW; 18-MAR-98.  
DR PDB, 1VMO; 18-MAR-98.  
DR PDB, 1VMP; 18-MAR-98.  
DR PDB, 1VMQ; 18-MAR-98.  
DR PDB, 1VMR; 18-MAR-98.  
DR PDB, 1SMF; 29-APR-98.  
DR PDB, 21ZA; 16-SEP-98.  
DR PDB, 21ZB; 16-SEP-98.  
DR PDB, 21ZC; 16-SEP-98.  
DR PDB, 21ZD; 23-SEP-98.  
DR PDB, 21ZE; 16-SEP-98.  
DR PDB, 21ZF; 16-SEP-98.  
DR PDB, 21ZG; 16-SEP-98.  
DR PDB, 21ZH; 16-SEP-98.  
DR PDB, 21ZI; 16-SEP-98.  
DR PDB, 21ZJ; 16-SEP-98.  
DR PDB, 21ZK; 16-SEP-98.  
DR PDB, 21ZL; 16-SEP-98.  
DR PDB, 2RTA; 18-NOV-98.  
DR PDB, 2RTB; 18-NOV-98.  
DR PDB, 2RTC; 18-NOV-98.  
DR PDB, 2RTD; 18-NOV-98.  
DR PDB, 2RTE; 18-NOV-98.  
DR PDB, 2RTE; 27-JAN-99.  
DR PDB, 2RTG; 18-NOV-98.  
DR PDB, 2RTH; 18-NOV-98.  
DR PDB, 2RTJ; 25-NOV-98.  
DR PDB, 2RTK; 18-NOV-98.  
DR PDB, 2RTL; 18-NOV-98.  
DR PDB, 2RTM; 18-NOV-98.  
DR PDB, 2RTN; 18-NOV-98.  
DR PDB, 2RTO; 18-NOV-98.  
DR PDB, 2RTP; 18-NOV-98.  
DR PDB, 2RTP; 18-NOV-98.  
DR PDB, 2RTR; 18-NOV-98.  
DR INTERPRO: IPR000088; -.  
DR PRAM: PF01382; Avidin; 1.  
DR PROSITE: PS00577; AVIDIN; 1.  
KW Signal; Biotin; 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 183 183 STREPTAVIDIN.  
FT BINDING 67 67 INVOLVED IN BIOTIN BINDING.  
FT BINDING 78 78 INVOLVED IN BIOTIN BINDING.  
FT BINDING 116 116 INVOLVED IN BIOTIN BINDING.  
FT BINDING 132 132 INVOLVED IN BIOTIN BINDING.  
FT BINDING 144 144 INVOLVED IN BIOTIN BINDING.  
FT SITE 83 85 CELL ATTACHMENT SITE.  
FT HELIX 38 41  
FT STRAND 43 47  
FT TURN 48 49  
FT STRAND 52 57  
FT TURN 59 60  
FT STRAND 62 68  
FT STRAND 77 84  
FT STRAND 95 104  
FT STRAND 109 121  
FT STRAND 127 136  
FT HELIX 140 145  
FT STRAND 147 156  
SQ SEQUENCE 183 AA; 18834 MW; 4652D8AE018468F5 CRC64;

Query Match 100.0%; Score 183; DB 1; Length 183;  
Best Local Similarity 100.0%; Pred. No. 1.5e-177;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRRIVAAIVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQLGSTFIVTAGAD 60
Db 1 MRRIVAAIVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQLGSTFIVTAGAD 60
QY 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTMSGQY 120
Db 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTMSGQY 120
QY 121 VGGAEARINQWLLTSTTANNAKSTLVGHDPFTTKPSPASISDAKKGAVNNGNLDA 180
Db 121 VGGAEARINQWLLTSTTANNAKSTLVGHDPFTTKPSPASISDAKKGAVNNGNLDA 180
QY 181 VQG 183
Db 181 VQG 183

RESULT 2
SAV2_STRVL STANDARD; PRT; 183 AA.
AC 053532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95359204; Pubmed=7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S78777; AAB35015.1; -.
DR HSSP: P22629; IPTS.
DR INTERPRO: IPR000088; -.
DR PFM: PF01382; Avidin; 1.
DR PROSITE: PS00577; AVIDIN; 1.
KW Signal; Biotin.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 183 STREPTAVIDIN V1.
FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 78 78 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
SQ SEQUENCE 183 AA; 18864 MW; 4652D8BB14847DE0 CRC64;
```

```
Best Local Similarity 100.0%; Pred. No. 6.4e-117;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRIVAAIVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQLGSTFIVTAGAD 60
Db 1 MRRIVAAIVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQLGSTFIVTAGAD 60
QY 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTMSGQY 120
Db 61 GALTGTYESAVGNAESRYVLGTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTMSGQY 120
QY 121 VQG 123
Db 121 VQG 123

RESULT 3
SAV2_STRVL STANDARD; PRT; 183 AA.
AC 053533;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE STREPTAVIDIN V2 PRECURSOR (SA V2).
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95359204; Pubmed=7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S78782; AAB35016.1; -.
DR HSSP: P22629; IPTS.
DR INTERPRO: IPR000088; -.
DR PFM: PF01382; Avidin; 1.
DR PROSITE: PS00577; AVIDIN; 1.
KW Signal; Biotin.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 183 STREPTAVIDIN V2.
FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 78 78 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
FT SIMILARITY).
SQ SEQUENCE 183 AA; 18833 MW; FEARFFDEA4ECCA CRC64;
```

Query Match 32.2%; Score 59; DB 1; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.1e-52;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY      63  LTGTESAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYRNAHSATWGGQYV 121
      |||||||
DB      63  LTGTESAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYRNAHSATWGGQYV 121

RESULT  4
NMR_NEUCR ID NMR_NEUCR STANDARD: PRT: 488 AA.
AC P23762;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NITROGEN METABOLIC REGULATION PROTEIN (NMR PROTEIN).
GN NMR.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=91042412; PubMed=2146484;
RA Young J.L., Jatal G., Fu Y.H., Marzluf G.A.;
RT "Nucleotide sequence and analysis of NMR, a negative-acting
RL regulatory gene in the nitrogen circuit of Neurospora crassa.";
RN Mol. Genet. 222:120-128(1990).
RN [2]
RP SEQUENCE OF 99-310 FROM N.A.
RX MEDLINE=92035098; PubMed=1834354;
RA Jatal G., Marzluf G.A.;
RT "Generation of new mutants of nmr, the negative-acting nitrogen
RL regulatory gene of Neurospora crassa, by repeat induced mutation.";
RN Curr. Genet. 20:283-288(1991).
RN [3]
RN MUTAGENESIS.
RP STRAIN=74-OR23-1A;
RX MEDLINE=91109706; PubMed=2148799;
RA Jatal G., Marzluf G.A.;
RT "Analysis of conventional and in vitro generated mutants of nmr, the
RL negatively acting nitrogen regulatory gene of Neurospora crassa.";
RN Mol. Genet. 222:233-240(1990).
CC -I- FUNCTION: MAJOR NEGATIVE REGULATORY PROTEIN IN THE NITROGEN
CC CONTROL CIRCUIT OF NEUROSPORA CRASSA WHICH ALLOWS UTILIZATION
CC OF VARIOUS NITROGEN-CONTAINING COMPOUNDS.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S64286; CAB32318.1; -.
CC PIR: S11910; S11910.
CC Transcription regulation; Repressor; Nitrate assimilation;
CC Nuclear protein.
CC KW MUTAGEN 2
CC FT MUTAGEN 386 P->K: LOSS OF FUNCTION.
CC FT MUTAGEN 386 G->D,R: LOSS OF FUNCTION.
CC FT MUTAGEN 386 G->C,S: FEW, VERY SLOWLY GROWING
CC FT TRANSFORMANTS.
CC FT SIMILAR 96 193 TO YEAST ARGRTII (AA 588-681).
CC FT DOMAIN 412 488 DISPENSABLE FOR NMR FUNCTION.
CC FT SEQUENCE 488 AA; 54857 MW; DDDAF91671C7846C CRC64;
SQ

Query Match 4.4%; Score 8; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT  5
TP2B_CRILLO ID TP2B_CRILLO STANDARD: PRT: 1612 AA.
AC Q64399;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN TOP2B.
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=96085121; PubMed=7495861;
RA Deraudre S., Frey S., Delaporte C., Jacquemin-Sablon A.;
RT "Cloning and characterization of full-length cDNAs coding for the DNA
RT topoisomerase II beta from Chinese hamster lung cells sensitive and
RL resistant 9-OH-ellipticine.";
RN Biochim. Biophys. Acta 1264:178-182(1995).
CC -I- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -I- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -I- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X86455; CA60173.1; -.
CC DR HSSP: P06786; IBGW.
CC DR INTERPRO: IPR000947; -.
CC DR INTERPRO: IPR001154; -.
CC DR INTERPRO: IPR001241; -.
CC DR INTERPRO: IPR002205; -.
CC DR PFAM: PF00204; DNA_topoisomII; 1.
CC DR PFAM: PF00521; DNA_topoisomI; 1.
CC DR PRINTS: PR00418; TP12FAMILY.
CC DR PRINTS: PR00615; CCAATSUBUNTA.
CC DR PRINTS: PR01158; TOPISMRASEII.
CC DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
CC KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
CC FT NP_BIND 170 175 ATP (POTENTIAL).
CC FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
CC FT SEQUENCE 1612 AA; 182074 MW; C01D6FC40620FC68 CRC64;
SQ

Query Match 4.4%; Score 8; DB 1; Length 1612;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      165 DAAKAGV 172
      |||||||
DB      161 DAAKAGV 188

```

```

RESULT  6
VLXS_LAMB

```



```

ID VLVS_LAMB      STANDARD:      PRT:      107 AA.
AC P03705;
DT 21-JUL-1966 (Rel. 01, Created)
DT 21-JUL-1966 (Rel. 01, Last sequence update)
DE HOLIN (GPS PROTEIN) [INCLUDES: LYSIS PROTEIN S; LYSIS INHIBITOR].
GN S.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA."
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=86304148; PubMed=2943725;
RA Raab R., Neal G., Garrett J., Grimalia R., Fusselman R., Young R.;
RT "Mutational analysis of bacteriophage lambda lysis gene S."
RL J. Bacteriol. 167:1035-1042(1986).
CC -1- FUNCTION: THE 105 AA PRODUCT INDUCES HOST CELL LYSIS BY CAUSING
CC FORMATION OF PORES IN THE INNER MEMBRANE. THE 107 AA PROTEIN
CC COUNTERACTS THE FORMATION OF SUCH TUNNELS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, INNER MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: USING ALTERNATIVE INITIATION CODONS IN
CC THE SAME READING FRAME, THE GENE TRANSLATES INTO TWO PRODUCTS:
CC THE LYSIS PROTEIN S AND AN INHIBITOR OF LYSIS.
CC -1- SIMILARITY: BELONGS TO THE LAMBDA PHAGE S PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J02459; AAA96597.1; -
DR EMBL: M14035; AAA32248.1; -
DR PIR: A04327; YVBPL.
KM Phage lysis protein; Transmembrane; Inner membrane;
KW Alternative Initiation.
FT CHAIN 1 107 LYSIS INHIBITOR.
FT INIT_MET 3 3 FOR LYSIS PROTEIN S.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
SQ SEQUENCE 107 AA; 11520 MW; 66D0D62426F1766E CRC64;

Query Match 3.8%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 AAKKAGV 172
DB 95 AAKKAGV 101

RESULT 7
VLT1_BPP22 STANDARD: PRT; 108 AA.
AC P09962;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE HOLIN (LYSIS PROTEIN 13).
GN 13.
OS Bacteriophage P22.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.

```

```

RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=86045883; PubMed=2998005;
RA Remell D., Potete A.R.;
RT "Phage P22 lysis genes: nucleotide sequences and functional
RT relationships with t4 and lambda genes."
RL Virology 143:280-289(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Schicklmaier P., Huber J., Schmeleger H.;
RL Submitted (Dec-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX Kropinski A.M.B., VanderByl C.S.;
RT "The completed sequence of genome of Salmonella phage P22."
RL Submitted (Dec-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR LYSIS OF THE BACTERIAL CELL WALL BY
CC DISRUPTING THE CELL MEMBRANE, THEREBY GIVING HYDROLYTIC ENZYMES
CC ACCESS TO THE CELL WALL.
CC -1- SIMILARITY: BELONGS TO THE LAMBDA PHAGE S PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M10997; AAA2265.1; -
DR EMBL: X67137; CAA47616.1; -
DR EMBL: AF217253; AAF75039.1; -
DR PIR: S22903; YVBPS2.
KM Phage lysis protein; Late protein.
SQ SEQUENCE 108 AA; 11682 MW; 6233F931AE7C787E CRC64;

Query Match 3.8%; Score 7; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 AAKKAGV 172
DB 96 AAKKAGV 102

RESULT 8
RS8_MICLU STANDARD: PRT; 132 AA.
AC P33106;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S8.
GN RPSH.
OS Micrococcus luteus (Micrococcus lysodeikticus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Micrococcaceae; Micrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90133967; PubMed=2533272;
RA Ohama T., Muto A., Osawa S.;
RT "Spectinomycin operon of Micrococcus luteus: evolutionary
RT implications of organization and novel codon usage."
RL J. Mol. Evol. 29:381-395(1989).
CC -1- FUNCTION: BINDS DIRECTLY TO THE CENTRAL DOMAIN OF 16S RIBOSOMAL
CC RNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X17524; CAA35561.1; -.  
 DR PIR: S29885; S29885.  
 DR HSSP: P56209; 1SET.  
 DR INTERPRO: IPR000630; -.  
 DR PFAM: PF00410; Ribosomal\_S8; 1.  
 DR PROSITE: PS00053; RIBOSOMAL\_S8; 1.  
 DR Ribosomal protein; rRNA-binding.  
 KW INIT\_MET 0 BY SIMILARITY.  
 SQ SEQUENCE 132 AA; 14363 MW; 0EDB952F12C3143D CRC64;

Query Match 3.8%; Score 7; DB 1; Length 132;  
 Best local Similarity 100.0%; Pred. No. 8.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 AAKAGV 172  
 Db 117 AAKAGV 123

RESULT 9  
 SOD4\_MAIZE STANDARD; PRT; 151 AA.  
 AC P23345;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE SUPEROXIDE DISMUTASE [CU-ZN] 4A (EC 1.15.1.1).  
 GN SODCC.3 OR SOD4P.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90136495; PubMed=2482436;  
 RA Cannon R.E., Scandalios J.G.;  
 RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in maize."  
 RL Mol. Gen. Genet. 219:1-8(1989).  
 CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
 CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
 CC -!- SUBUNIT: HOMODIMER.  
 CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
 CC -!- SUBUNIT: HOMODIMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- MISCELLANEOUS: EUKARYOTIC CELLS CONTAIN A MITOCHONDRIAL MN-CONTAINING ENZYME & A CYTOPLASMIC CU-ZN-CONTAINING ENZYME.  
 CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

FT METAL 44 44 COPPER (BY SIMILARITY).  
 FT METAL 46 46 COPPER (BY SIMILARITY).  
 FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).  
 FT METAL 69 69 ZINC (BY SIMILARITY).  
 FT METAL 78 78 ZINC (BY SIMILARITY).  
 FT METAL 81 81 ZINC (BY SIMILARITY).  
 FT METAL 118 118 COPPER (BY SIMILARITY).  
 FT DISULFID 55 144 BY SIMILARITY.  
 SQ SEQUENCE 151 AA; 14983 MW; 9C6226F86C919E58 CRC64;

Query Match 3.8%; Score 7; DB 1; Length 151;  
 Best local Similarity 100.0%; Pred. No. 9.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 VTAGADG 61  
 Db 85 VTAGADG 91

RESULT 10  
 SOD5\_MAIZE STANDARD; PRT; 151 AA.  
 AC P23346;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE SUPEROXIDE DISMUTASE [CU-ZN] 4AP (EC 1.15.1.1).  
 GN SODCC.2 OR SOD4AP.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90136495; PubMed=2482436;  
 RA Cannon R.E., Scandalios J.G.;  
 RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in maize."  
 RL Mol. Gen. Genet. 219:1-8(1989).  
 CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
 CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
 CC -!- SUBUNIT: HOMODIMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- MISCELLANEOUS: EUKARYOTIC CELLS CONTAIN A MITOCHONDRIAL MN-CONTAINING ENZYME & A CYTOPLASMIC CU-ZN-CONTAINING ENZYME.  
 CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

FT METAL 78 78 ZINC (BY SIMILARITY).  
 FT METAL 81 81 ZINC (BY SIMILARITY).  
 FT METAL 118 118 COPPER (BY SIMILARITY).  
 FT DISULFID 55 144 BY SIMILARITY.  
 SO SEQUENCE 151 AA; 14939 MW; 9C7E572A6C1AEF1D CRC64;

Query Match 3.8%; Score 7; DB 1; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 9.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VTGADG 61  
 |||||  
 DB 85 VTGADG 91

RESULT 11  
 SODC\_CAVPO STANDARD; PRT; 152 AA.  
 AC P33431;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE SUPEROXIDE DISMUTASE [CU-ZN] (EC 1.15.1.1).  
 GN SOD1.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 RN Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HARTLEY; TISSUE-LUNG;  
 RA Yuan H.T., Bingle C.D., Kelly F.J.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE:  
 RC TISSUE-LIVER;  
 MEDLINE=94059414; PubMed=8240718;  
 RA Wolf B., Relnecke K., Aumann K.-D., Brigelius-Flohe R., Flohe L.;  
 RT "Taxonomical classification of the guinea pig based on its Cu/Zn  
 superoxide dismutase sequence.";  
 RL Biol. Chem. Hoppe-Seyler 374:641-649(1993).  
 CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
 CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
 CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- MISCELLANEOUS: EUKARYOTIC CELLS CONTAIN A MITOCHONDRIAL  
 MN-CONTAINING ENZYME & A CYTOPLASMIC CU-ZN-CONTAINING ENZYME.  
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U39844; AAC52720.1; -.  
 DR PIR: S36108;  
 DR HSP: P00441; 1SPD.  
 DR INTERPRO: IPR001424; -.  
 DR PFAM: PF00080; sodbu.1.  
 DR PRINTS: PRO0068; CUZNDISMUTASE.  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_1; 1.  
 DR PROSITE: PS00332; SOD\_CU\_ZN\_2; 1.  
 KM Oxidoreductase; Copper; Zinc; Acetylation.  
 FT INIT\_MET 0 0  
 FT MOD\_RES 1 1 ACETYLATION.  
 FT METAL 45 45 COPPER (BY SIMILARITY).  
 FT METAL 47 47 COPPER (BY SIMILARITY).  
 FT METAL 62 62 COPPER AND ZINC (BY SIMILARITY).  
 FT METAL 70 70 ZINC (BY SIMILARITY).

FT METAL 79 79 ZINC (BY SIMILARITY).  
 FT METAL 82 82 ZINC (BY SIMILARITY).  
 FT METAL 119 119 COPPER (BY SIMILARITY).  
 FT DISULFID 56 145 BY SIMILARITY.  
 FT CONFLICT 102 103 LI -> IL (IN REF. 2).  
 SO SEQUENCE 152 AA; 15538 MW; 5891CE16E2ACBFB3 CRC64;

Query Match 3.8%; Score 7; DB 1; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 9.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VTGADG 61  
 |||||  
 DB 86 VTGADG 92

RESULT 12  
 CBPM\_STRAL STANDARD; PRT; 255 AA.  
 AC P00733;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE MURAMOYL-PENTAPEPTIDE CARBOXYPEPTIDASE PRECURSOR (EC 3.4.17.8)  
 DE (METALLO DD-PEPTIDASE) (ZN DD-PEPTIDASE) (D-ALANYL-D-ALANINE  
 DE CARBOXYPEPTIDASE).  
 OS Streptomyces albus G.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 RN Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-G;  
 RA MEDLINE=91115060; PubMed=2276609;  
 RA Duez C., Lakaye B., Houba S., Dusart J., Ghuyssen J.-M.;  
 RT "Cloning, nucleotide sequence and amplified expression of the gene  
 RT encoding the extracellular metallo (Zn) DD-peptidase of Streptomyces  
 RT albus G.";  
 RL FEMS Microbiol. Lett. 59:215-219(1990).  
 RN [2]  
 RP SEQUENCE OF 43-255.  
 RC STRAIN-SOLVIFACIENS;  
 RX MEDLINE=83131648; PubMed=6825689;  
 RA Joris B., van Beumen J., Casagrande F., Gerday C., Frere J.-M.,  
 RA Ghuyssen J.-M.;  
 RT "The complete amino acid sequence of the Zn<sup>2+</sup>-containing D-alanyl-D-  
 RT alanine-cleaving carboxypeptidase of streptomyces albus G.";  
 RL Eur. J. Biochem. 130:53-69(1983).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=83012968; PubMed=7121586;  
 RA Dideberg O., Charlier P., Dive G., Joris B., Frere J.-M.,  
 RA Ghuyssen J.-M.;  
 RT "Structure of a Zn<sup>2+</sup>-containing D-alanyl-D-alanine-cleaving  
 RT carboxypeptidase at 2.5-A resolution.";  
 RL Nature 299:469-470(1982).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RA Wery J.-P., Charlier P., Dideberg O.;  
 RL Submitted (MAR-1996) to the PDB data bank.  
 CC -1- FUNCTION: THIS ENZYME CATALYZES CARBOXYPEPTIDATION AND  
 TRANSEPTIDATION REACTIONS INVOLVED IN BACTERIAL CELL WALL  
 METABOLISM. IT EFFECTIVELY CATALYZES THE TRANSFER OF THE N-ALPHA,  
 CC N-EPSILON-DIACETYL-L-LYSYL-D-ALANYL ELECTROPHILIC GROUP OF THE  
 CC STANDARD TRIPEPTIDE SUBSTRATE N-ALPHA,N-EPSILON-DIACETYL-L-LYSYL-  
 CC D-ALANYL-D-ALANINE TO WATER. IT ALSO PERFORMS A WEAK BETA-  
 CC LACTAMASE ACTIVITY, HYDROLYSING PENICILLIN INTO PENICILLATE AT A  
 CC VERY LOW RATE.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES THE C-TERMINAL D-ALANYL-D-ALANINE  
 CC BOND IN PEPTIDOGLYCAN.  
 CC -1- COFACTOR: ZINC.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- PTM: THE N-TERMINUS IS PARTIALLY BLOCKED AS A RESULT OF THE

```

CC CYCLIZATION OF THE FIRST TWO AA INTO ANHYDROSPARTYLGLYCINE IMIDE.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X55794; CAA39319.1; -.
CC PIR: A00913; CPSPMU.
CC PDB: 1LBV; 08-NOV-96.
CC MEROPS: M15.001; -.
CC INTERPRO: IPR002477; -.
CC PFAM: PF01471; PG_binding_1.1.
CC Hydrolase; Carboxypeptidase; Cell wall; Zinc; 3D-structure; Signal.
CC SIGNAL 1 42
CC CHAIN 43 255
CC MOD_RES 43 43
CC DISULFID 45 123
CC DISULFID 136 184
CC DISULFID 212 253
CC METAL 196 196
CC METAL 237 237
CC METAL 239 239
CC BINDING 180 180
CC ACT_SITE 234 234
CC CONFLICT 43 43
CC CONFLICT 110 110
CC SIMILAR 43 109
CC SO SEQUENCE 255 AA; 26190 MW; 9168C2145A863AD3 CRC64;
    TO B. SUBTILIS AUTOLYSIN (AA 194-261).

Query Match 3.8%; Score 7; DB 1; Length 255;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TASASAD 25
Db 32 TASASAD 38

RESULT 13
EIL_RAT
ID EIL_RAT STANDARD; PRT; 266 AA.
AC P00773;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).
GN ELA1
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Pictet R.L.,
RA Nikovits W., Rutter W.J.;
RT "Primary structure of two distinct rat pancreatic preproelastases
RT determined by sequence analysis of the complete cloned messenger
RT ribonucleic acid sequences."
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Ibane R.G.,
RA Rutter W.J., MacDonald R.J.;
RT "Structure of the two related elastase genes expressed in the rat
RT pancreas."
RL J. Biol. Chem. 259:14271-14278(1984).

```

```

RN [3]
RP SEQUENCE OF 17-45.
RC TISSUE-PANCREAS;
RX MEDLINE=84000385; PubMed=6555050;
RA Larrman C.;
RT Isolation and characterization of rat pancreatic elastase.";
RL Biochemistry 22:3763-3770(1983).
CC -1 FUNCTION: ACTS UPON ELASTIN.
CC -1 CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.
CC PREFERENTIAL CLEAVAGE: ALA-I-XNA.
CC -1 SUBCELLULAR LOCATION: SECRETED.
CC -1 TISSUE SPECIFICITY: PANCREAS.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V01234; CAA24544.1; -.
CC EMBL: L00117; AAA98811.1; -.
CC EMBL: L00112; AAA98811.1; JOINED.
CC EMBL: L00113; AAA98811.1; JOINED.
CC EMBL: L00114; AAA98811.1; JOINED.
CC EMBL: L00115; AAA98811.1; JOINED.
CC EMBL: L00116; AAA98811.1; JOINED.
CC PIR: A00960; ELRT1.
CC HSSP: P00772; 1EST.
CC MEROPS: S01.153; -.
CC INTERPRO: IPR001254; -.
CC INTERPRO: IPR001314; -.
CC PFAM: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; Pancreas; Zymogen; Signal.
CC SIGNAL 1 16
CC PROPEP 17 26
CC CHAIN 27 266
CC DISULFID 56 72
CC DISULFID 153 220
CC DISULFID 184 200
CC DISULFID 210 240
CC ACT_SITE 71 71
CC ACT_SITE 119 119
CC ACT_SITE 214 214
CC CONFLICT 104 104
CC CONFLICT 108 108
CC CONFLICT 244 244
CC CONFLICT 266 266
CC SO SEQUENCE 266 AA; 28976 MW; 5A56F8FCF1AEDA CRC64;

Query Match 3.8%; Score 7; DB 1; Length 266;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 VGGAER 127
Db 28 VGGAER 34

RESULT 14
DAPA_PROMA
ID DAPA_PROMA STANDARD; PRT; 302 AA.
AC P49423;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)

```

```

DE DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52) (DHDPs).
GN DAPA.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCMP 1375.
RA Lorenz M., Boerner T., Hess W.R.;
RT "Molecular cloning and characterization of a dihydrodipicolinate
RT synthase (DHDPs) gene from the photoautotrophic prokaryote
RT Prochlorococcus marinus CCMP 1375 (Prochlorophyta).";
RL Endocyt. Cell Res. 11:59-68(1995).
CC -1- CATALYTIC ACTIVITY: L-ASPARTATE 4-SEMIALDEHYDE + PYRUVATE =
CC DIHYDRODIPICOLINATE + 2 H(2)O.
CC -1- ENZYME REGULATION: SENSITIVE TO LYSINE INHIBITION
CC (BY SIMILARITY).
CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOSOLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DHDPs FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z68126; CAA92211.1; -.
CC HSP: P05640; IDHP.
CC INTERPRO: IPR001899; -.
CC DR INTERPRO: IPR002220; -.
CC DR PFAM: PF00701; DHDPs_1.
CC DR PRINTS: PR00146; DHPICSNTHASE.
CC DR PROSITE: PS00665; DHDPs_1; 1.
CC DR PROSITE: PS00666; DHDPs_2; 1.
CC KM Lyase; Diaminopimelate biosynthesis; lysine biosynthesis;
CC Feedback-inhibition.
CC FT ACT_SITE 172 172 BY SIMILARITY.
CC SQ SEQUENCE 302 AA; 31769 MW; F11F6023643FB84 CRC64;

Query Match 3.8%; Score 7; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AGADGAL 63
DB 105 AGADGAL 111

RESULT 15
PSD4_MOUSE
ID PSD4_MOUSE STANDARD; PRT; 376 AA.
AC 035226;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 26S PROTEASOME REGULATORY SUBUNIT 55A (MULTIUBQUITIN CHAIN BINDING
DE PROTEIN).
GN PSMD4 OR MCB1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98112730; PubMed-9511739;
RA Pusch W., Jaehner D., Ivell R.;
RT "Molecular cloning and testicular expression of the gene transcripts
RT encoding the murine multiubiquitin-chain-binding protein (Mcbl).";

```

```

RL Gene 207:19-24(1998).
CC -1- FUNCTION: BINDS AND PRESUMABLY SELECTS UBIQUITIN-CONJUGATES FOR
CC DESTRUCTION.
CC -1- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT 55A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AF013099; AAC53547.1; -.
CC DR MGD; MGI:1201670; PSMD4.
CC KW Proteasome.
CC SQ SEQUENCE 376 AA; 40703 MW; 732AC02B56760EAA CRC64;

Query Match 3.8%; Score 7; DB 1; Length 376;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SAAEAGI 41
DB 242 SAAEAGI 248

Search completed: April 6, 2001, 01:29:25
Job time: 399 sec

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 6, 2001, 01:12:46 ; Search time 67.82 Seconds  
(without alignments)  
316.264 Million cell updates/sec

Title: US-09-589-870-2  
Perfect score: 183  
Sequence: 1 MRKIVAAIAVSLFTVSITA.....IDAKKAGVNGNPLDAVQ 183

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_15:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	4.4	271	10	Q9M8S8
2	8	4.4	366	2	Q9Z8K0
3	8	4.4	366	2	Q9PKW4
4	8	4.4	366	2	Q9K274
5	8	4.4	366	2	Q9JSG3
6	8	4.4	416	2	Q84078
7	8	4.4	428	3	Q99006
8	8	4.4	470	2	Q9RTZ8
9	8	4.4	1067	2	Q9JY67
10	8	4.4	1067	2	Q9JY51
11	7	3.8	86	2	Q9WVF8
12	7	3.8	86	2	Q9RBR3
13	7	3.8	86	2	Q9K363
14	7	3.8	106	9	Q9MGN5
15	7	3.8	107	9	Q9MCT3
16	7	3.8	117	10	Q9SL91
17	7	3.8	124	10	P93800
18	7	3.8	149	2	Q9ZCK5
19	7	3.8	149	2	O05444

20	7	3.8	152	10	P93801	P93801 zea mays (m
21	7	3.8	164	2	Q9Z582	Q9Z582 streptomyc
22	7	3.8	166	1	Q9YCD6	Q9YCD6 aeropyrum p
23	7	3.8	182	2	Q9KSV2	Q9KSV2 vibrio chol
24	7	3.8	191	1	Q9WMQ3	Q9WMQ3 pyrococcus
25	7	3.8	204	10	Q9LVA6	Q9LVA6 arabidopsis
26	7	3.8	218	2	Q51920	Q51920 prochloroco
27	7	3.8	225	5	Q61752	Q61752 caenorhabd
28	7	3.8	251	2	Q9K3Y4	Q9K3Y4 streptomyc
29	7	3.8	255	10	Q9W9D1	Q9W9D1 arabidopsis
30	7	3.8	260	11	Q9JWM0	Q9JWM0 mus musculu
31	7	3.8	268	4	Q9NS92	Q9NS92 homo sapien
32	7	3.8	275	1	Q54485	Q54485 staphylothe
33	7	3.8	295	10	Q22158	Q22158 arabidopsis
34	7	3.8	298	2	Q9RM95	Q9RM95 bradyrhizob
35	7	3.8	300	2	Q52615	Q52615 proteus vul
36	7	3.8	319	2	P96845	P96845 mycobacteri
37	7	3.8	324	2	Q9Z178	Q9Z178 lactococcus
38	7	3.8	325	10	Q9ZRC5	Q9ZRC5 glycine max
39	7	3.8	332	2	Q87803	Q87803 pseudomonas
40	7	3.8	345	10	Q9R024	Q9R024 arabidopsis
41	7	3.8	346	10	Q9SQH3	Q9SQH3 brassica na
42	7	3.8	349	11	Q9JWM1	Q9JWM1 mus musculu
43	7	3.8	354	2	Q9KW67	Q9KW67 staphylococ
44	7	3.8	368	11	Q9JWM2	Q9JWM2 mus musculu
45	7	3.8	379	11	Q9JWM3	Q9JWM3 mus musculu

## ALIGNMENTS

### RESULT 1

Q9M8S8 ID Q9M8S8 PRELIMINARY: PRT: 271 AA.  
AC Q9M8S8;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE PUTATIVE MTO-IMOSITOL MONOPHOSPHATASE.  
GN F13E7.19.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RT "Arabidopsis thaliana chromosome III BAC F13E7 genomic sequence.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC018363; AAF26973.1; -  
SQ SEQUENCE 271 AA: 29121 MW: 2785848F541B8F5F CRC64;

Query Match 4.4%; Score 8; DB 10; Length 271;

Best local Similarity 100.0%; Pred. No. 8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 IDAKKAG 171  
Db 15 IDAKKAG 22

RESULT 2  
Q9Z8K0 PRELIMINARY: PRT: 366 AA.  
AC Q9Z8K0;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE DNA POLYMERASE III (BETA CHAIN).  
 GN DNAN.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Greenwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 DR EMBL: AE001618; AAD18487.1; -.  
 DR INTERPRO: IPR001001; -.  
 DR PRAM: PR00712; DNA\_pol3\_beta: 1.  
 SQ SEQUENCE 366 AA; 40334 MW; 8A6D188DB26BC5E1 CRC64;

Query Match 4.4%; Score 8; DB 2; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82  
 |||||  
 Db 149 ESRVLTG 156

RESULT 3  
 O9PKW4 PRELIMINARY; PRT; 366 AA.  
 AC O9PKW4;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE DNA POLYMERASE III, BETA SUBUNIT.  
 GN TC0347.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MOPN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39";  
 RN Nucleic Acids Res. 28:1397-1406(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,  
 RA Hickey E.K., Peterson J., Umayam L.A., Uterback T., Berry K.,  
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 RA Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G.,  
 RA Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWI029 from USA";  
 DR TIGR: TC0347; -.  
 DR INTERPRO: IPR001001; -.  
 DR PRAM: PR00712; DNA\_pol3\_beta: 1.  
 SQ SEQUENCE 366 AA; 40493 MW; E3887C463C442D72 CRC64;

Db 149 ESRVLTG 156

RESULT 4  
 O9K274 PRELIMINARY; PRT; 366 AA.  
 AC O9K274;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE DNA POLYMERASE III, BETA SUBUNIT.  
 GN CP0419.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39";  
 RN Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AE002203; AAF38262.1; -.  
 DR TIGR: CP0419; -.  
 SQ SEQUENCE 366 AA; 40440 MW; EC54D84CFA085150 CRC64;

Query Match 4.4%; Score 8; DB 2; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82  
 |||||  
 Db 149 ESRVLTG 156

RESULT 5  
 O9JSG3 PRELIMINARY; PRT; 366 AA.  
 AC O9JSG3;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE DNA POLYMERASE III (BETA CHAIN).  
 GN DNAN.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii R., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWI029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL: AP002546; BAA96548.1; -.  
 DR PRAM: PR00712; DNA\_pol3\_beta: 1.  
 SQ SEQUENCE 366 AA; 40364 MW; E8FA92ACF76BD1B0 CRC64;

Query Match 4.4%; Score 8; DB 2; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 ESRVLTG 82  
 |||||  
 Db 149 ESRVLTG 156

Query Match 4.4%; Score 8; DB 2; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

75 ESRVLTG 82  
 |||||



```

RESULT 6
OB4078      PRELIMINARY;      PRT;      416 AA.
AC 084078;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE DNA POL III (BETA CHAIN).
GN DNAN.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL; AE001282; AAC67666.1; -.
DR HSSP; P00583; 2POL.
DR INTERPRO; IPR001001; -.
DR PRAM; PF00712; DNA_pol3_beta; 1.
SQ SEQUENCE 416 AA; 46529 MW; F328D087A8930683 CRC64;

```

```

Query Match      4.4%; Score 8; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 75 ESRVLTG 82
Db 199 ESRVLTG 206

```

```

RESULT 7
Q99006      PRELIMINARY;      PRT;      428 AA.
AC 099006;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ENDOCHITINASE.
OS Trichoderma hamatum.
OC Eukaryota; Fungi; Ascomycota; anamorphic Ascomycota; Trichoderma.
OX NCBI_Taxid=49224;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAM-61;
RX MEDLINE=97132641; PubMed=8978093;
RA Pekete C., Weszely T., Hornok L.;
RT "Assignment of a PCR-amplified chitinase sequence cloned from
RT Trichoderma hamatum to resolved chromosomes of potential biocontrol
RT species of Trichoderma."
RL FEMS Microbiol. Lett. 145:385-391(1996).
DR EMBL; Z71415; CAA96021.1; -.
DR HSSP; P07254; 1CTN.
DR INTERPRO; IPR001223; -.
DR INTERPRO; IPR001579; -.
DR PRAM; PF00704; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KM Hydrolase; Glycosidase.
SQ SEQUENCE 428 AA; 46881 MW; 0901641A987D8A59 CRC64;

```

```

Query Match      4.4%; Score 8; DB 3; Length 428;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 159 PSASIDA 166
Db 139 PSASIDA 146

```

```

RESULT 8
Q9RT28      PRELIMINARY;      PRT;      470 AA.
AC 09RT28;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN DRI599.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_Taxid=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the Radioresistant Bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
DR EMBL; AE002004; AAF11162.1; -.
DR TIGR; DRI599; -.
DR INTERPRO; IPR000126; -.
DR PRINTS; PR00839; V8PROTEASE.
SQ SEQUENCE 470 AA; 50683 MW; 60D8581386287562 CRC64;

```

```

Query Match      4.4%; Score 8; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 57 AGADGALT 64
Db 392 AGADGALT 399

```

```

RESULT 9
Q9JY67      PRELIMINARY;      PRT;      1067 AA.
AC 09JY67;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE MULTIPLE TRANSFERABLE RESISTANCE SYSTEM PROTEIN MTRD.
GN NMB1715.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2015755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Debey R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cline H., Vamathevan J.,
RA Cotton M.D., Uitterback T.R., Khouli H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Piza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain

```

RT MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL: AF002521; AAF42062.1; -  
 DR TIGR: NMB1715; -  
 SQ SEQUENCE 1067 AA; 113973 MW; A05545371B7943C2 CRC64;

Query Match  
 Best Local Similarity 4.4%; Score 8; DB 2; Length 1067;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAEAGITG 43  
 |||||||  
 Db 1054 AAEAGITG 1061

RESULT 10  
 O9JUT51  
 ID O9JUT51 PRELIMINARY; PRT; 1067 AA.

AC O9JUT51;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DE PROBABLE DRUG EFFLUX PROTEIN.  
 GN MTRD OR NMA1969.

OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 NCBI\_TaxID=65699;

RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,  
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;

RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491.";  
 RL Nature 404:502-506(2000).  
 DR EMBL: AL162757; CAB85189.1; -

SQ SEQUENCE 1067 AA; 114012 MW; D3216DD7F4783B41 CRC64;

Query Match  
 Best Local Similarity 4.4%; Score 8; DB 2; Length 1067;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAEAGITG 43  
 |||||||  
 Db 1054 AAEAGITG 1061

RESULT 11

O9WMP8  
 ID O9WMP8 PRELIMINARY; PRT; 86 AA.

AC O9WMP8;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE HRCS.

OS Xanthomonas campestris (pv. vesicatoria).  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 NCBI\_TaxID=341;

OX NCB1\_TaxID=341;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=75-3;  
 RX MEDLINE=98453137; PubMed=9781876;

RA Huguet E., Hahn K., Wengelnik K., Bonas U.;  
 RT "DppA mutants of Xanthomonas campestris pv. vesicatoria are affected

RT in pathogenicity but retain the ability to induce host-specific  
 RT hypersensitive reaction.";  
 RL Mol. Microbiol. 29:1379-1390(1998).  
 DR EMBL: AF056246; AAD21322.1; -  
 DR INTERPRO: IPR002191; -  
 DR PFAM: PF01313; Bac\_export\_3; 1.  
 DR PRINTS: PR00952; TYPE3IMOPROT.  
 SQ SEQUENCE 86 AA; 8914 MW; CD44BF70739BAD82 CRC64;

Query Match  
 Best Local Similarity 3.8%; Score 7; DB 2; Length 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 VVAIAIV 11  
 |||||||  
 Db 58 VVAIAIV 64

RESULT 12  
 O9RBR3  
 ID O9RBR3 PRELIMINARY; PRT; 86 AA.

AC O9RBR3;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE HRCS HOMOLOG.

OS Xanthomonas campestris (pv. glycines).  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 NCBI\_TaxID=36865;

RP SEQUENCE FROM N.A.  
 RC STRAIN=8RA;  
 RA Oh C., Heu S., Yoo J.Y., Cho Y.S.;

RT "A hcrv-homologous gene mutant of Xanthomonas campestris pv. glycines  
 RT 8ra that lost pathogenicity on the host plant but was able to elicit  
 RT the hypersensitive response on nonhosts.";  
 RL Mol. Plant Microbe Interact. 12:633-639(1999).

DR EMBL: AF160974; AAD46903.1; -  
 DR INTERPRO: IPR002191; -  
 DR PFAM: PF01313; Bac\_export\_3; 1.  
 DR PRINTS: PR00952; TYPE3IMOPROT.

SQ SEQUENCE 86 AA; 8928 MW; DA4FD2D1A39BAD82 CRC64;

Query Match  
 Best Local Similarity 3.8%; Score 7; DB 2; Length 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 VVAIAIV 11  
 |||||||  
 Db 58 VVAIAIV 64

RESULT 13

O9K363  
 ID O9K363 PRELIMINARY; PRT; 86 AA.

AC O9K363;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE HRCS (HRP3).

OS Xanthomonas oryzae pv. oryzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 NCBI\_TaxID=64187;

OX NCB1\_TaxID=64187;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=MAFP 311018;  
 RA Ochiai H.;  
 RT "Construction and characterization of a Xanthomonas oryzae pv. oryzae

RT bacterial artificial chromosome library."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF301237;  
 RA Oku T.;  
 RT "Clustered hrp genes in Xanthomonas oryzae pv. oryzae."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB045311; BAB07863.1; -;  
 DR EMBL: AB040134; BAA92827.1; -;  
 SQ SEQUENCE 86 AA; 8944 MW; 3844AA65739BAD98 CRC64;

Query Match 3.8%; Score 7; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VVAIAV 11  
 DB 58 VVAIAV 64

RESULT 14  
 ID Q9MCN5 PRELIMINARY; PRT; 106 AA.  
 AC Q9MCN5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HOLIN.  
 GN 70.  
 OS Bacteriophage HK97.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;  
 OC Lambda phage group.  
 OX NCBI\_TaxID=37554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Juhala R., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,  
 RT "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic  
 RT mosaicism in the lambdaoid bacteriophages."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF069529; AAF31096.1; -;  
 SQ SEQUENCE 106 AA; 11274 MW; CA91992E1C30769E CRC64;

Query Match 3.8%; Score 7; DB 9; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAKKAGV 172  
 DB 94 AAKKAGV 100

RESULT 15  
 ID Q9MCT3 PRELIMINARY; PRT; 107 AA.  
 AC Q9MCT3;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HOLIN.  
 OS Bacteriophage HK022.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;  
 OC Lambda phage group.  
 OX NCBI\_TaxID=10742;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Juhala R., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,  
 RT "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic  
 RT mosaicism in the lambdaoid bacteriophages."  
 SQ SEQUENCE 107 AA; 11340 MW; 18A92A2A20FDICAB CRC64;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF069308; AAF30386.1; -;  
 SQ SEQUENCE 107 AA; 11340 MW; 18A92A2A20FDICAB CRC64;

Query Match 3.8%; Score 7; DB 9; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAKKAGV 172  
 DB 95 AAKKAGV 101

Search completed: April 6, 2001, 01:28:06  
 Job time: 920 sec

1



vector (pBluescript II SK+) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=56) and sequence (n=16). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 65 a 78 c 83 g 38 t  
ORIGIN

alignment\_scores: Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-589-870-2 x A1973896 ..

Align seg 1/1 to: A1973896 from: 1 to: 264

166 AAlaAlaLysLysAlaGlyValAsnAsn 174  
|||||  
167 GCGGCCAGAGAGCGCGGTGACACAC 193

seq\_name: gb\_gss14:A0998089

seq\_documentation\_block:

LOCUS A0998089 439 bp DNA GSS 24-FEB-2000  
DEFINITION RPCI-23-279B21.TV RPCI-23 Mus musculus genomic clone RPCI-23-279B21  
, DNA sequence.

ACCESSION A0998089  
VERSION A0998089.1 GI:7073186

KEYWORDS GSS:  
house mouse.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 439)

AUTHORS Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akkurat,  
, B., Levin, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C. M.

TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
MEDLINE Other GSSs: RPCI-23-279B21.TV

COMMENT Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.bufileo.edu). Clones may be purchased from  
BACpac Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)  
or from Resear h Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/cdb/bac.ends/mouse/bac\_end\_intro.html  
Plate: 279 row: B column: 21  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..439

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-279B21"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 152 a 84 c 104 g 99 t  
ORIGIN

alignment\_scores: Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-589-870-2 x A0998089 ..

Align seg 1/1 to: A0998089 from: 1 to: 439

75 GLuserArgTyrrValIleuThrGlyArg 83  
|||||  
158 GAAGACAGGTATGCTCCTGAGGAGA 184

seq\_name: gb\_gss12:A0836896

seq\_documentation\_block:

LOCUS A0836896 574 bp DNA GSS 30-AUG-1999  
DEFINITION HS\_5463\_A1\_H09.T7A RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate-1039 Col-17 Row=O, DNA sequence.

ACCESSION A0836896  
VERSION A0836896.1 GI:5806770

KEYWORDS GSS:  
human.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 574)

AUTHORS Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and  
Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.bufileo.edu). Clones may be purchased from  
BACpac Resources (http://bacpac.med.bufileo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 1039 row: O column: 17  
Seq primer: T7  
Class: BAC ends

FEATURES  
source Location/Qualifiers  
1..574

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-1039 Col-17 Row=O"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"  
BASE COUNT 220 a 87 c 109 g 149 t 9 others  
ORIGIN

```

Alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

Alignment_block:
US-09-589-870-2 x AQ836896/rev ..

Align seg 1/1 to reverse of: AQ836896 from: 1 to: 574

      10 AlaValSerLeuThrThyValSerIle 18
      |||||
185 GCTGTCAGCTCCTCAGTACTGTCGTATT 159

seq_name: gb_esc175:BE726886

seq_documentation_block:
LOCUS      BE726886      606 bp      mRNA      EST      14-SEP-2000
DEFINITION      894096B10.y2 C. reinhardtii CC-1690, normalized, lambda Zap II
ACCESSION      Chlamydomonas reinhardtii cDNA, mRNA sequence.
VERSION      BE726886
KEYWORDS      BE726886.1 GI:10128182
SOURCE      EST.
ORGANISM      Chlamydomonas reinhardtii.
                Chlamydomonas reinhardtii
                Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
                Chlamydomonadaceae; Chlamydomonas.
REFERENCE      1 (bases 1 to 606)
AUTHORS      Groseman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
                McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model,
                Unicellular System for Analyzing Gene Function and Regulation in
                Vascular Plants. Project phase 2
                Unpublished (2000)
COMMENT      Contact: Charles Hauser
                DCMB Box 91000
                Duke University
                Durham, NC 27708-1000
                Tel: 919 613 8159
                Fax: 919 613 8177
                Email: chauser@duke.edu.
FEATURES
     source              Location/Qualifiers
     1..606
     /organism="Chlamydomonas reinhardtii"
     /strain="CC-1690 wild type mt+ 21gr"
     /db_xref="taxon:3055"
     /clone_id="C. reinhardtii CC-1690, normalized, lambda Zap
     II"
     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
     XhoI; This library, constructed by John Davies and Jeffrey
     Mcdermott, combines cDNAs from CC-1690 cells grown to
     mid-log phase in TAP (acetate-containing) medium in the
     light, TAP medium in the dark, HS (minimal) medium in
     ambient levels of CO2 and HS medium bubbled with 5% CO2.
     PolyA mRNA was purified from each sample, pooled and cDNA
     synthesized. The cDNA was directionally cloned into lambda
     Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
     pBluescript II SK- plasmids were excised from the lambda
     Zap clones by superinfection with Exs85sst (Stratagene)
     phage. The library was normalized using method 4 described
     in Bonardo et al (1996) Genome Research 6: 791-806."
BASE COUNT      108 a      180 c      214 g      104 t
ORIGIN

```

```

US-09-589-870-2 x BE726886  ..
Align seg 1/1  to: BE726886  from: 1  to: 606

164  llaaspaaialaalsylsylaaglyval 172
|||||
468  ATCGATGCTGCCAGACAGAGCGGGCGTC 494

seq_name: gb_est11:AA749932

seq_documentation_block:
LOCUS      AA749932      184 bp      mRNA      EST      20-JAN-1998
DEFINITION  ISAS0492. Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION  AA749932
VERSION    AA749932.1  GI:2796638
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
1 (bases 1 to 184)
Nhm,B.H., Kim,J.K., Cheong,S.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee
,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
Contact: Eun M. Y.
Department of Cytogenetics
National Inst. of Agril. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel.: 82 331 290 0301
Fax: 82 331 290 0307
Email: mweunsun20.ast.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea, 449-728 dhnam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..184
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="ISAS0492"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/notes="Vector: pluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
BASE COUNT      31 a      79 c      29 g      45 t
ORIGIN

FEATURES
source
1..184
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="ISAS0492"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/notes="Vector: pluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."

Alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

Alignment_block:
US-09-589-870-2 x AA749932  ..

Align seg 1/1  to: AA749932  from: 1  to: 184

20  AlaseraAlaseraAlasproser 27
|||||
69  GCCTCGCCTCGCGCGATCATCC 92

seq_name: gb_est50:AM894265

seq_documentation_block:
LOCUS      AM894265      223 bp      mRNA      EST      24-MAY-2000
DEFINITION  CM3-NN0030-110500-178-f11 NN0030 Homo sapiens cDNA, mRNA sequence.

```

ACCESSION AM894265  
 VERSION AM894265.1 GI:8058470  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 223)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ft2-CM3-NN0030-110>)  
 500-178-flt4t3=2000-05-11&ft4=1)  
 Seg primer: puc 18 forward  
 High quality sequence start: 22  
 High quality sequence stop: 77.  
 Location/Qualifiers  
 1..223  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NN0030"  
 /dev\_stage="Adult"  
 /note="Organ: nervous, normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 low stringency conditions."  
 BASE COUNT 62 a 53 c 39 g 69 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-589-870-2 x AM894265 ..  
 Align seg 1/1 to: AM894265 from: 1 to: 223  
 seq\_name: gb\_est16:AI084542  
 seq\_documentation\_block: 258 bp mRNA EST 27-AUG-1998  
 LOCUS AI084542  
 DEFINITION ov43004.x1 Soares, testis\_NHT Homo sapiens cDNA clone IMAGE:1640071  
 3', mRNA sequence.  
 ACCESSION AI084542  
 VERSION AI084542.1 GI:3422965  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 258)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo Ph.D.  
 ' cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
 Insert length: 900 Std Error: 0.00  
 Seg primer: -40m3 fwd. ET from Amersham  
 High quality sequence stop: 221.  
 Location/Qualifiers  
 1..258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1640071"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories 'Inc' and primed with a Not I - oligo(dT) primer [5' TGTTCACATCTGAAGTGGAGCGGCCGCCCATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 62 a 73 c 77 g 46 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-589-870-2 x AI084542/rev ..  
 Align seg 1/1 to reverse of: AI084542 from: 1 to: 258  
 seq\_name: gb\_est24:AI756574  
 seq\_documentation\_block: 299 bp mRNA EST 18-JAN-2000  
 LOCUS AI756574  
 DEFINITION Eteshea02f10.y1 Elmeria M5-6 Merozoite stage Elmeria tenella cDNA 5', mRNA sequence.  
 ACCESSION AI756574  
 VERSION AI756574.1 GI:5150297  
 KEYWORDS EST.  
 SOURCE Elmeria tenella.  
 ORGANISM Elmeria tenella  
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriidae; Elmeriidae;  
 Elmeria.  
 1 (bases 1 to 299)  
 Liberators, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,



TITLE  
JOURNAL  
COMMENT  
Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Riltner, E., Koh, S., Florence, N., Shin, T., Jackson, Y., Cardenas, M., Mcann, R., Waterston, R., Wilson, R. and Sibley, D. WashU-Merck Elmeria tenella project  
Unpublished (1999)  
Contact: David Sibley, Ph.D.  
WashU-Merck Elmeria tenella project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estevenson.wustl.edu  
Contact David Sibley (toxeist@orcim.wustl.edu) for further information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Glbco  
High quality sequence stop: 298.  
Location/Qualifiers

## FEATURES

source

1. .299  
/organism="Elmeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Elmeria M5-6 Merozoite stage"  
/dev\_stage="Merozoite"  
/lab\_host="SOLR E. coli"  
/note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI ; Merozoites were obtained from ceecal scrapings of chickens infected with E. tenella. The library may contain a small percentage of host or bacterial contaminants. cDNA was synthesized from poly RNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on Sephacryl S500. cDNAs were digested with EcoRI/XhoI and cloned into lambda Zap II (Stratagene). Clones were converted to phagemids by mass excision using Exassist helper phage and SOLR cells (Stratagene) . Insert sizes range from 0.7-1.5 kb."  
BASE COUNT 38 a 103 c 84 g 73 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x A1756574/rev ..

Align seg 1/1 to reverse of: A1756574 from: 1 to: 299

32 AlaGlnValSerAlaAlaGluAla 39  
|||||  
77 GCACAGGTTCTGCTGCTGAGCGT 54  
seq\_name: gb\_est17:A1194189

seq\_documentation\_block:  
LOCUS A1194189 311 bp mRNA EST 13-OCT-1998  
DEFINITION ue81c04.r1 Soares\_NMPu Mus musculus cDNA clone IMAGE:1497510 5',  
RNA sequence.  
ACCESSION A1194189  
VERSION A1194189.1 GI:3745396  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 311)  
Marfa, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

TITLE  
JOURNAL  
COMMENT  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:935114  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 213.  
Location/Qualifiers

## FEATURES

source

1. .311  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:1497510"  
/clone\_lib="Soares\_NMPu"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pregnant mouse uterus, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 86 a 52 c 62 g 111 t  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x A1194189 ..

Align seg 1/1 to: A1194189 from: 1 to: 311

12 SerLeuThrThrValSerIleThr 19  
|||||  
115 TCACCTACACTGTCAGTATTACC 138  
seq\_name: gb\_est37:AV540058

seq\_documentation\_block:  
LOCUS AV540058 326 bp mRNA EST 07-SEP-2000  
DEFINITION AV540058 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
CDNA clone R2143h06f 3', mRNA sequence.  
ACCESSION AV540058  
VERSION AV540058.1 GI:8701816  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

REFERENCE  
AUTHORS  
TITLE  
Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 326)  
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries  
DNA Res. 7, 175-180 (2000)  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Erika Asamizu

The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: [asamazu@kazusa.or.jp](mailto:asamazu@kazusa.or.jp), URL: <http://www.kazusa.or.jp/en/plant/location/Qualifiers>

## FEATURES

### source

```

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="R214h05F"
/clone_id="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/issue="Vector: plasmid pIII SK; site_1: EcoRI; site_2:
XhoI"

```

BASE COUNT	93 a	67 c	72 g	94 t
ORIGIN				

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000
```

```
alignment_block:
```

US-09-589-870-2 X AV540058

Align seg 1/1 to: AV540058 from: 1 to: 326

```

164 1leasplatalalysrlysalagly 171
      |||||
31  ATTGATGCCGCTAAAAAAGCTGGA 54
seq_name: gb_est53:BB059043

```

seq\_name: gb\_est53:BB059043

```
seq_documentation_block:
```

seq. documentation:	333 bp	EST	01-AUG-2000
LOCUS	BB059043	mRNA	
DEFINITION	BB059043 RIKEN full-length enriched, 2 days neonate sympathetic ganglion Mus musculus cDNA clone 7120486x15 3', mRNA sequence.		

ACCESSION	BB059043
VERSION	BB059043.1
	GI:8466191

SOURCE ORGANISM	house mouse. Mus musculus
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

CONGRATULATIONS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
1 (bases 1 to 333)  
REFERENCE  
komno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
AUTHORS

Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, I., Itoh, M.  
 Pitoze, T., Fukuda, S., Fukunishi, T., Goto, A., Hayashi, K.,  
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kuribara, C., Kusabe  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki  
 , Ono, T., Owa, C., Saito, H., Sakai, C., Sato, C., Shibata, K., Shiba  
 , Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara  
 , Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., To  
 , Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamataka  
 , Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yosh  
 , Muraetsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Kono, H., et al.)  
 Unpublished (2000)  
 Title  
 Journal  
 Comment  
 Contact: Yoshihide Hayashizaki  
 RIKEN Mouse ESTs Group, Life Science Tsukuba Center.

**FEATURES**  
source

trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

Itom, M., Kitsuana, T., Akiyama, J., Shibata, K., Iwawa, M., Kawai, J., Tomaru, Y., Carinci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carinci, P. and Hayashizaki, Y. High efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-46 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```

FEATURES
source
Location/Qualifiers
1. 333
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7120466K15"
/clone_lib="RIKEN full-length enriched, 2 days neonate
sympathetic ganglion"
sex="mixed"
/tissue_type="sympathetic ganglion"
/dev_stage="2 days neonate"
/lab_host="DH10B"
/notes="Site.1: SalI; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCCAAGAGCTCTTTTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGAGATTCGACGATTTATTTAAATTAATCCCGCCCGCCG 3']. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda PflC I.
Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
~RNA was provided by Akira Nakagawara, Div. of
Biochemistry, Chiba Cancer Research Institute,
666-2 Nitona, Chuoh-ku, Chiba, 260-8717 Japan, whose
assistance we gratefully acknowledge."

```

```
alignment_scores:      Length: 8
      quality:      8.00
      ratio:      1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
```

alignment\_block:  
no 00-500-870-2 x PR059043

US-09-589-870-2 X BB059043 .

Align seg 1/1 to: BB059043 from: 1 to: 333

```

146 SerThrLeuValGlyHisaspThr 153
      |||||
141 AGCACTTAGTGGACATGATACT 164

```

seq\_name: gb\_est2:AA126930

```
seq_documentation_block:
```

seq_documentation_block:			
LOCUS	AA126930	373 bp	EST 26-NOV-1996
DEFINITION	z123s01.r1 Soares_pregnant_uterus_NbHPV Homo sapiens cDNA clone		
IMAGE:	502728 5', mRNA sequence.		

ACCESSION	AA126930	
VERSION	AA126930.1	GI:1687951
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE 1 (bases 1 to 373)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 TITLE JOURNAL  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watsn.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 295.  
 Location/Qualifiers  
 source 1..373  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3807716"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:502728"  
 /clone\_1lb="Soares\_pregnant\_uterus\_NbHPV"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pT73-Pac; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', AACCTGAGAGATTCGCGCGCCCTTTTCTTTTCTTTT 3'1', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."  
 BASE COUNT 87 a 89 c 98 g 94 t 5 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-589-870-2 x AA126930 ..  
 Align seg 1/1 to: AA126930 from: 1 to: 373  
 131 GlnTrpleuthrSerglyThr 138  
 ||||||||||||||||||  
 320 CAGTGCCTCTGACTTCAGGACN 343  
 seq\_name: gb\_gss23:B40058  
 seq\_documentation\_block:  
 LOCUS B40058 374 bp DNA GSS 18-OCT-1997  
 DEFINITION HS-1050-B2-F08-WF-ab1 CIT Human Genomic Sperm Library C Homo  
 sapiens genomic clone Plate=CT 772 Col=16 Row=L, DNA sequence.  
 ACCESSION B40058  
 VERSION B40058.1 GI:2544310  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 374)  
 AUTHORS Mahairas, G.G., Zackrone, K.D., Smith, T., Tipson, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.  
 TITLE Construction of a Characterized Clone Resource for Genomic

Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors  
 Unpublished (1997)  
 COMMENT Contact: Mahairas GC, Zackrone KD, Hood L  
 University of Washington  
 Seattle, WA 98195, USA  
 Tel: (206) 616-8744  
 Fax: (206) 685-7301  
 Email: kzackrone@u.washington.edu  
 Sequence Tagged Connector  
 Plate: CT 772 row: L column: 16  
 Class: BAC ends  
 High quality sequence stop: 374.  
 Location/Qualifiers  
 source 1..374  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=CT 772 Col=16 Row=L"  
 /clone\_1lb="CIT Human Genomic Sperm Library C"  
 /sex="M"  
 /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in E-Coli DH10B"  
 BASE COUNT 111 a 66 c 78 g 119 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-589-870-2 x B40058/rev ..  
 Align seg 1/1 to reverse of: B40058 from: 1 to: 374  
 133 LeuethrSerglyThrTrnGlu 140  
 ||||||||||||||||||  
 368 CTACTCAGTCAGGACACCGAA 345  
 seq\_name: gb\_gss16:AZ131983  
 seq\_documentation\_block:  
 LOCUS AZ131983 377 bp DNA GSS 02-JUN-2000  
 DEFINITION OSJNB0111P08r CUGI Rice BAC Library (ECORI) Oryza sativa genomic  
 clone OSJNB0111P08r, DNA sequence.  
 ACCESSION AZ131983  
 VERSION AZ131983.1 GI:8210478  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
 REFERENCE 1 (bases 1 to 377)  
 AUTHORS Wing, R.A. and Dean, R.A.  
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Seq primer: CGAAGACGCTATGACCATG  
 Class: BAC ends  
 High quality sequence start: 14  
 High quality sequence stop: 363.  
 Location/Qualifiers  
 source 1..377  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"

```

/db.xref="taxon:4530"
/clone_1ib="CUC1 Rice BAC Library (EcoRI)"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

BASE COUNT 87 a 71 c 82 g 136 t 1 others

ORIGIN

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

```

alignment_block:
US-09-589-870-2 x AZ131983/rev ..
```

```

Align seg 1/1 to reverse of: AZ131983 from: 1 to: 377
```

```

14 ThrThrValSerIleThrAlaSer 21
|||||
360 ACCACCGTGTCTATTACGCGCAGC 337
```

```
seq_name: gb_estf2:BE553360
```

```
seq_documentation_block:
```

```

LOCUS BE553360 390 bp mRNA 15-AUG-2000
DEFINITION ur45c10.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153234 5',
mRNA sequence.
```

```

ACCESSION BE553360
VERSION BE553360.1 GI:9817847
KEYWORDS EST.
```

```
SOURCE house mouse.
```

```
ORGANISM Mus musculus
```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE 1 (bases 1 to 390)
```

```
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

```
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
```

```
JOURNAL Unpublished (1997)
```

```
COMMENT Contact: Robert Strausberg, Ph.D.
```

```
Tel: (301) 496-1550
```

```
Email: Robert_Strausberg@nih.gov
```

```
This clone is available royalty-free through LNL; contact the
```

```
IMAGE Consortium (info@image.lnl.gov) for further information.
```

```
MGI:1055990
```

```
Seq primer: -40RP from Gibco
```

```
High quality sequence stop: 310.
```

```
FEATURES
  Location/Qualifiers
```

```
1..390
  /organism="Mus musculus"
```

```

/strain="FVB-3"
/db.xref="taxon:10090"
/clone_1ib="IMAGE:3153234"
/clone_1ib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
```

BASE COUNT 110 a 68 c 87 g 125 t

ORIGIN

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

```

alignment_block:
US-09-589-870-2 x BE553360 ..
```

```
Align seg 1/1 to: BE553360 from: 1 to: 390
```

```

12 SerLeuThrThrValSerIleThr 19
|||||
356 TCACTTACAACTGTCTATTACG 379
```

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2001, 19:57:40 ; Search time 73.6 Seconds  
(without alignments)  
85.020 Million cell updates/sec

Title: US-09-589-870-2  
Perfect score: 936  
Sequence: 1 MRKIVAAIAIVSLTTSITFA.....IDAAKAGVNGNPLDAVQQ 183

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq-36:\*

- 1: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1980.DAT:\*
- 2: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1981.DAT:\*
- 3: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1982.DAT:\*
- 4: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1983.DAT:\*
- 5: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1984.DAT:\*
- 6: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1985.DAT:\*
- 7: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1986.DAT:\*
- 8: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1987.DAT:\*
- 9: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1988.DAT:\*
- 10: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1989.DAT:\*
- 11: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1990.DAT:\*
- 12: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1991.DAT:\*
- 13: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1992.DAT:\*
- 14: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1993.DAT:\*
- 15: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1994.DAT:\*
- 16: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1995.DAT:\*
- 17: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1996.DAT:\*
- 18: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1997.DAT:\*
- 19: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1998.DAT:\*
- 20: /cgn2\_2/gcgdata/geneseq/geneseqp/AA1999.DAT:\*
- 21: /cgn2\_2/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	936	100.0	183	7 P60625	Sequence of a stre
2	936	100.0	183	10 P93530	Streptavidin prote
3	936	100.0	183	14 R44491	Streptavidin gene.
4	936	100.0	183	18 W29306	Wild-type streptav
5	936	100.0	183	19 W59216	S. avidinii strept
6	936	100.0	183	20 Y17868	Streptococcus stre
7	936	100.0	183	21 Y44701	Streptavidin prote
8	936	100.0	186	21 Y80512	Streptomyces avidi
9	931	99.5	184	21 Y80513	Streptomyces avidi
10	928	99.1	182	8 P70492	Streptavidin sequ
11	928	99.1	183	21 Y84020	Amino acid sequenc
12	919	98.2	183	19 W59217	S. avidinii strept

13	917	98.0	183	19 W59218	S. avidinii strept
14	901.5	96.3	186	21 Y84021	Amino acid sequenc
15	899.5	96.1	185	21 Y84022	Amino acid sequenc
16	846	90.4	435	15 R56483	ScFV pRAS108 and p
17	838	89.5	161	21 Y80515	Streptomyces avidi
18	835	89.2	162	21 Y80514	Streptomyces avidi
19	835	89.2	163	9 P80160	Biosynthetic prote
20	834	89.1	160	10 P93531	Mature streptavidin
21	829	88.6	159	18 W29310	Streptavidin prote
22	829	88.6	159	18 W29311	Streptavidin prote
23	828	88.5	159	18 W29309	Streptavidin prote
24	824	88.0	159	18 W29314	Streptavidin prote
25	823	87.9	159	18 W29313	Streptavidin prote
26	822	87.8	159	18 W29316	Streptavidin prote
27	821	87.7	159	18 W29315	Streptavidin prote
28	820	87.6	159	18 W29312	Streptavidin prote
29	818	87.4	159	18 W29319	Streptavidin prote
30	816	87.2	159	18 W29318	Streptavidin prote
31	815	87.1	159	18 W29317	Streptavidin prote
32	810	86.5	159	18 W29320	Streptavidin prote
33	745	79.6	415	15 R56484	ScFV pRAS109 and p
34	686	73.3	270	20 Y28928	S. avidinii protein
35	679	72.5	168	21 Y44700	Potato proteinase
36	679	72.5	673	17 W04208	Streptavidin/lucif
37	675	72.1	128	14 R34722	Core streptavidin.
38	675	72.1	128	17 W04211	Streptomyces avidi
39	675	72.1	128	18 W29308	Recombinant Core-s
40	673.5	72.0	402	15 R56485	ScFV pRAS110 and p
41	660	70.5	685	17 W04209	Mutant streptavidin
42	656	70.1	140	17 W04210	Mutant Streptomyces
43	650	69.4	122	21 Y80516	Streptomyces avidi
44	650	69.4	122	21 Y80517	Streptomyces avidi
45	203	21.7	40	8 P70491	N-terminal sequenc

#### ALIGNMENTS

RESULT	1
P60625	
ID	P60625 standard; Protein: 183 AA.
XX	
AC	P60625;
XX	
DT	13-AUG-1991 (first entry)
XX	
DE	Sequence of a streptavidin-like polypeptide encoded by SA307.
XX	
KW	Antibiotic; biotin binding affinity; fusion protein.
XX	
OS	Streptomyces.
XX	
PN	W08602077-A.
XX	
PD	10-APR-1986.
XX	
PF	01-OCT-1985; 85WO-0001901.
XX	
PR	02-OCT-1984; 84US-0656873.
XX	
PA	(MEAD/) MEADE H M.
XX	
PI	Meade HM, Garvin JL, Biogen NV;
XX	
DR	WPI: 1986-106643/16.
XX	
DR	N-PSDB; N60626.
XX	
PT	DNA sequences and hybrid DNA sequences - encoding
PT	streptavidin-like polypeptide, also joined to another protein.
XX	e.g. tissue plasminogen activator
XX	
PS	Disclosure; Fig. 2; 54pp; English.
XX	

CC The inventors claim the DNA sequence in SA307 which codes for a  
CC streptavidin-like polypeptide (see N60626), and the polypeptide  
CC encoded by it (P60625). They also claim hybrid SOS comprising N60626  
CC and a second sequence coding for another protein, polypeptide,  
CC peptide or AA (pref. tissue plasminogen activator (tPA)).

XX Sequence 183 AA;

Query Match 100.0%; Score 936; DB 7; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2.5e-76;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIYVAIAIVSLITVSIPTASADPSKDSKAQVSAEAGITGWNQIGSTFIYTAGAD 60  
1 mrlivvaiaivslitvsiptasasdpkskdaqvsaeeagltgwnqigstfiytagad 60

Db 1 mrlivvaiaivslitvsiptasasdpkskdaqvsaeeagltgwnqigstfiytagad 60  
61 galtgyesavgaesryvltgrydsapaldgsqtaglgtvawknyrnahsattwsgqy 120

QY 61 GALTGYESAVGAESRYVLTGRYDSAPATDGSQTALGWTVMKNNYRNAHSATTWSGOY 120  
121 VGGAEARINTQWLTSGTTPANAMKSTLVGHDFPTKPKPSAASIDAARKAGVNNGNPLDA 180  
121 vggaearin tqwltsgtteanawksltvghdftrkpkpsaasidaarkagvnnpnlda 180

QY 181 VQO 183  
181 vqg 183

RESULT 2  
P93530  
ID P93530 standard; protein; 183 AA.

XX P93530;

DT 04-JUN-1990 (first entry)

DE Streptavidin protein.

XX Streptavidin; Streptomyces avidinii; biotin.

XX Streptomyces avidinii.

OS Streptomyces

XX Key Location/Qualifiers

FT Peptide 1..24 /note="Leader sequence"

FT Protein 25..159 /note="This sequence was as the basis for the design  
for the synthetic gene of the present invention."

XX MO8903422-A.

XX 20-APR-1989.

XX 07-OCT-1988; 88WO-GB00831.

XX 08-OCT-1987; 87GB-0023661.

XX (BRBI-) BRIT BIO-TECHN LTD.

XX Edwards RM;

XX WPI; 1989-130040/17.

XX DNA sequence encoding streptavidin and vector  
XX comprising hybrid gene encoding fusion protein with  
XX biotin-binding activity

XX Fig 1; page 1/5; 22pp; English.

CC Streptavidin is a 60KD protein isolated from Streptomyces avidinii that  
CC binds extremely tightly to the vitamin biotin. It is composed of four

CC identical subunits of 15KD and binds 4 mole of biotin per mole of  
CC protein. It is structurally related to the protein avidin. It can be  
CC readily conjugated to a range of other proteins. In order to facilitate  
CC the incorporation of streptavidin into expression vectors and the  
CC production of novel chimeric proteins containing streptavidin  
CC functionality, an improved novel synthetic gene for streptavidin has  
CC been constructed (n90755) based on the amino acid sequence of mature  
CC streptavidin.

XX Sequence 183 AA;

Query Match 100.0%; Score 936; DB 10; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2.5e-76;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIYVAIAIVSLITVSIPTASADPSKDSKAQVSAEAGITGWNQIGSTFIYTAGAD 60  
1 mrlivvaiaivslitvsiptasasdpkskdaqvsaeeagltgwnqigstfiytagad 60

Db 1 mrlivvaiaivslitvsiptasasdpkskdaqvsaeeagltgwnqigstfiytagad 60  
61 galtgyesavgaesryvltgrydsapaldgsqtaglgtvawknyrnahsattwsgqy 120

QY 61 GALTGYESAVGAESRYVLTGRYDSAPATDGSQTALGWTVMKNNYRNAHSATTWSGOY 120  
121 VGGAEARINTQWLTSGTTPANAMKSTLVGHDFPTKPKPSAASIDAARKAGVNNGNPLDA 180  
121 vggaearin tqwltsgtteanawksltvghdftrkpkpsaasidaarkagvnnpnlda 180

QY 181 VQO 183  
181 vqg 183

RESULT 3  
R44491  
ID R44491 standard; Protein; 183 AA.

XX R44491;

DT 27-JUN-1994 (first entry)

DE Streptavidin gene.

XX Streptavidin; protein secretion; Bacillus subtilis.

XX Streptomyces avidinii.

OS Streptomyces

XX Key Location/Qualifiers

FT Protein 1..24 /label="signal\_peptide"

FT Peptide 25..183 /label="streptavidin"

FT Misc-difference 1..24 /note="expressed by transformed B. subtilis"

XX WO9324631-A.

XX 09-DEC-1993.

XX 27-MAY-1993; 93WO-US05240.

XX 29-MAY-1992; 92US-0891524.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Nagatajan V;

XX WPI; 1993-405822/50.

XX P-PSDB; 053412.

PT Streptavidin prodn. from Bacillus subtilis - using signal protein  
from bacterial exo-protein and expression element from Gram



PT positive bacterial protein.  
XX  
XX Disclosure; Fig 1b; 54pp; English.  
XX  
XX Tetrameric biologically active streptavidin is produced by secretion  
CC from *Bacillus subtilis* transformed with a plasmid encoding the  
CC sequence.  
XX  
SQ Sequence 183 AA;  
  
Query Match 100.0%; Score 936; DB 14; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2.5e-76;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKRIYVAIAIVSLITVYITRISASADPSKDSKAQVSAEAGITGTWYQLGSTFIYTAGAD 60  
DB 1 mrkivvaiaivslitvysltasasadpskdskaqvsaeeagltgtwnqlgstflvtagad 60  
QY 61 GALTGYESAIVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120  
DB 61 galtgtyesavgnaesryvltgrydsapadgsctalgwtvwmknnyrnhsaltwsqgy 120  
QY 121 VGGAEARINTQWILTSGTTEANAMKSTLVGHDPFTTKVYKPSAASIDAAKKAGVNNGNPLDA 180  
DB 121 vggaearin tqwiltsgtteanawkstlvghdfttkvkpsaasidaakkagvnngnplda 180  
QY 181 VQO 183  
DB 181 vqg 183  
  
RESULT 4  
W29306 ID W29306 standard; Protein; 183 AA.  
XX  
AC W29306;  
XX  
DT 27-APR-1998 (first entry)  
XX  
DE Wild-type streptavidin protein.  
XX  
KM Streptavidin; biotin; anti-interference reagent; detection; mutain;  
XX avidin; non-specific binding.  
OS Unidentified.  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..24  
FT /label= signal  
FT Protein 25..183  
XX  
PN DE19637718-A1.  
XX  
PD 02-OCT-1997.  
XX  
PF 16-SEP-1996; 96DE-1037718.  
XX  
PR 01-APR-1996; 96DE-1013053.  
XX  
PA (BOEF ) BOEHRINGER MANNHEIM GMBH.  
XX Brandstetter H, Deger A, Engh R, Kopetzki E, Mueller R;  
PI Schmitt U;  
XX  
XX WPI; 1997-482043/45.  
XX N-PSDB; T73193.  
XX  
XX Streptavidin and avidin mutants with reduced binding affinity for  
PT biotin - useful for reducing interference from nonspecific binding  
PT in assays  
XX  
XX Disclosure; Page 17-18; 26pp; German.

XX This sequence represents a streptavidin which is used in a novel method  
CC of reducing interference from non-specific binding in assays. Mutains  
CC constructed from a core streptavidin or avidin sequence are selected that  
CC differ from the native polypeptide by at least one amino acid and have a  
CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable  
CC polypeptide may be present as a polymeric conjugate, e.g. with another  
CC polypeptide or protein, especially bovine serum albumin. These mutains  
CC are used as anti-interference reagents for reducing and/or avoiding  
CC nonspecific interactions in a process for detecting an analyte. In  
CC particular, they are used in assays where the streptavidin/avidin-biotin  
CC specific binding pair is involved for qualitative and/or quantitative  
CC determination of an analyte in a test sample, e.g. a heterogeneous  
CC immunoassay or a hybridisation assay. Despite having a lower binding  
CC affinity for biotin, the mutains have high immunological cross-reactivity  
CC with native streptavidin and avidin.  
XX  
SQ Sequence 183 AA;  
  
Query Match 100.0%; Score 936; DB 18; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2.5e-76;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKRIYVAIAIVSLITVYITRISASADPSKDSKAQVSAEAGITGTWYQLGSTFIYTAGAD 60  
DB 1 mrkivvaiaivslitvysltasasadpskdskaqvsaeeagltgtwnqlgstflvtagad 60  
QY 61 GALTGYESAIVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120  
DB 61 galtgtyesavgnaesryvltgrydsapadgsctalgwtvwmknnyrnhsaltwsqgy 120  
QY 121 VGGAEARINTQWILTSGTTEANAMKSTLVGHDPFTTKVYKPSAASIDAAKKAGVNNGNPLDA 180  
DB 121 vggaearin tqwiltsgtteanawkstlvghdfttkvkpsaasidaakkagvnngnplda 180  
QY 181 VQO 183  
DB 181 vqg 183  
  
RESULT 5  
W59216 ID W59216 standard; Protein; 183 AA.  
XX  
AC W59216;  
XX  
DT 27-AUG-1998 (first entry)  
XX  
DE S. avidinli streptavidin protein.  
XX  
KM Streptavidin; ligand; binding affinity; mutant; isolation;  
XX purification; recover; immobilise.  
OS Streptomycetes avidinli.  
XX  
PN EP835934-A2.  
XX  
PD 15-APR-1998.  
XX  
PF 09-OCT-1997; 97EP-0117504.  
XX  
PR 10-OCT-1996; 96DE-1041876.  
XX  
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.  
XX Skerra A, Voss S;  
PI  
XX  
XX WPI; 1998-218868/20.  
XX N-PSDB; V34714.  
XX  
XX Streptavidin mutants with higher binding affinity for peptide  
PT ligands - have mutation in amino acid region 44-53, used to isolate,  
PT

PT purify or determine fusion proteins including these ligands  
XX  
PS Disclosure; Page -: 21pp; German.  
XX  
XX This sequence encodes a wild-type streptavidin protein isolated from  
CC Streptomyces avidinii. This sequence is used to produce mutants which  
CC are used in a method to assay the binding affinity of streptavidin  
CC mutants. These mutants have a mutation within the amino acid (aa) region  
CC 44-53 of the wild-type protein show a higher binding affinity than the  
CC wild-type for peptide ligands that include the sequence of formula  
CC  $\text{trp-x-his-pro-gln-phe-y-z}$  where x = any aa; y and z are both gly,  
CC or y = glu and z = arg or lys. Recombinant streptavidin mutants can  
CC be used to isolate, purify and determine proteins or to determine/recover  
CC substances that contain streptavidin-binding groups. Such compounds may  
CC also be used to immobilise fusions on microtitre plates, microbeads or  
CC sensor chips.  
CC NOTE: This sequence does not appear in the specification but is used to  
CC make the mutant streptavidin proteins represented in W59217 and W59218.  
XX  
XX Sequence 183 AA:

Query Match 100.0%; Score 936; DB 19; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2.5e-76;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRIIVAAIVSLTITVSIPTASASADPSKDSKAQVSAAEAGITGTWNOLGSTRITVAGAD 60  
DB 1 mrlivvaalavsltlvtlsasadspskdskaqvsaeeagltgwnqlgstlvtvagaad 60  
QY 61 GALTGYESAVGNAESRYVLTGRYDSAPATDGSGTALGTVANKNNYRNAHSATVWSGOY 120  
DB 61 galtgyesavgnaesryvltgrydsapatdgsqgtalgvawknnyrnhsatvwsqgy 120  
QY 121 VGGAERINTOWLLTSGTPEANAMKSTLVGHDTFTVKRPSAASIDAARKAGVNGNPLDA 180  
DB 121 vggaearinltwlltsgtpeanmkstlvghdtftvk rpsaasidaarkagvngnplda 180  
QY 181 VQG 183  
DB 181 vqg 183

RESULT 6  
Y17868  
ID Y17868 standard; Protein: 183 AA.  
XX  
AC Y17868;  
XX  
DT 20-AUG-1999 (first entry)  
XX  
DE Streptococcus streptavidin.  
XX  
XX Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;  
KW hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin.  
XX  
OS Streptococcus sp.  
XX  
PN WO9929838-A1.  
XX  
PD 17-JUN-1999.  
XX  
PF 09-DEC-1998; 98WO-US26086.  
XX  
PR 09-DEC-1997; 97US-0067978.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Cederholm-Williams SA;  
XX  
DR WPI; 1999-385599/32.  
XX  
XX N-PSDB; X80198.

PT A fibrinogen-converting enzyme fusion protein  
XX  
PS Disclosure; Page 28; 35pp; English.  
XX  
XX

CC The present invention describes a fibrinogen-converting enzyme fusion  
CC protein (FCE). The fusion protein is a multidomain protein comprising:  
CC (a) a FCE; and (b) a first member of a binding pair, that is linked to  
CC the FCE chain: (i) directly by bonds utilizing the N-terminal amino  
CC groups, the C-terminal carboxy groups or side-chain functionalities;  
CC (ii) via a bifunctional linkage moiety linking the groups or  
CC functionalities; or (iii) by the first member binding to the second  
CC member of the binding pair, where the second member of the binding pair  
CC is covalently attached to the first polypeptide chain. The FCE can be  
CC used in a method for producing fibrin. Fibrin is useful as a sealant in  
CC surgery to, e.g. reduce bleeding by sealing blood vessels, and tissues  
CC that have been dissected either in surgery or through wounding. The  
CC fusion protein allows for the removal of the fibrinogen converting  
CC enzyme from the fibrin sealant preparation via the binding of  
CC streptavidin to a biotin solid support. The present sequence represents  
CC Streptococcus streptavidin as given in the present invention.  
XX  
XX Sequence 183 AA:

Query Match 100.0%; Score 936; DB 20; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2.5e-76;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRIIVAAIVSLTITVSIPTASASADPSKDSKAQVSAAEAGITGTWNOLGSTRITVAGAD 60  
DB 1 mrlivvaalavsltlvtlsasadspskdskaqvsaeeagltgwnqlgstlvtvagaad 60  
QY 61 GALTGYESAVGNAESRYVLTGRYDSAPATDGSGTALGTVANKNNYRNAHSATVWSGOY 120  
DB 61 galtgyesavgnaesryvltgrydsapatdgsqgtalgvawknnyrnhsatvwsqgy 120  
QY 121 VGGAERINTOWLLTSGTPEANAMKSTLVGHDTFTVKRPSAASIDAARKAGVNGNPLDA 180  
DB 121 vggaearinltwlltsgtpeanmkstlvghdtftvk rpsaasidaarkagvngnplda 180  
QY 181 VQG 183  
DB 181 vqg 183

RESULT 7  
Y44701  
ID Y44701 standard; Protein: 183 AA.  
XX  
AC Y44701;  
XX  
DT 25-APR-2000 (first entry)  
XX  
DE Streptavidin protein for recombinant PART27 vector.  
XX  
XX  
KW Potato proteinase inhibitor-II; PPI-II; streptavidin; worm;  
KW insect; plant-noxious protein; pest resistance; moth; insect; weevil;  
KW grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;  
KW insecticidal.  
XX  
OS Unidentified  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..24  
FT /note= "Signal peptide"  
XX  
PN WO200004049-A1.  
XX  
PD 27-JAN-2000.  
XX  
PF 15-JUL-1999; 99WO-NZ00110.  
XX  
XX 15-JUL-1998; 98NZ-0331002.

XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.  
PA Christeller JF, Sutherland PM, Murray C, Markwick NP, Phillip BA;  
PI Malone LA, Burgess EPJ;  
XX  
XX WPI: 2000-171244/15.  
DR N-PSDB: 249867.  
XX  
XX New chimeric polypeptide and composition comprising the polypeptide  
PT useful for conferring pest resistance on plants -  
PI  
XX  
XX Disclosure: Fig 12; 11pp; English.  
XX  
XX The present sequence is streptavidin, a plant-toxicous protein.  
CC Recombinant vector, PART7 expressing a chimeric polypeptide comprising  
CC streptavidin mature peptide fused to the potato proteinase inhibitor-II  
CC (PPI-II) signal peptide is targeted to the vacuole.  
CC Transformation of plant genome with the vector can produce pest  
CC resistance in plants, plant derived products and stored harvest  
CC material. Pests that can be controlled include: cotton bollworm,  
CC tropical army-worm, European corn borer or red mite, tobacco horn worm,  
CC loopers, rice stem borer, portia, cutworms, diamondback moth, potato  
CC tuber moth, codling moth, indian meal moth, gypsy moth, argentine stem  
CC weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat  
CC weevils, mealworms, flour beetles, black field cricket, locusts,  
CC sawflies, Western flower thrips, Hessian flies or two-spotted mite.  
XX  
SQ Sequence 183 AA:

Query Match 100.0%; Score 936; DB 21; Length 183;

Best Local Similarity 100.0%; Pred. No. 2.5e-76;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRIIVAAIAVSLTTVSITNASADPSKDSKAQVSAEAGITGTWYNQSGTFTVTAGAD 60  
DB 1 mklivvaalavslitvtasasadpskdskaqvsaaegitgtwngqstftvttagad 60  
OY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120  
DB 61 galgtgyesaavgnaesryvltgrysdapatdgsctalgtwvmaknnyrnhsattwsqgy 120  
OY 121 VGGAERINTQWLLTSGTTEANAMKSTLVGHDTFTYVKPSAASIDAARKAGVNNGNPLDA 180  
DB 121 vggaearrintqwilltsgtteanawkstlvghdtftkvkpsaasidaarkagvnngnplda 180  
OY 181 YQQ 183  
DB 181 yqq 183

RESULT 8

Y80512 Y80512 standard; Protein: 186 AA.

AC Y80512;

DT 06-JUN-2000 (first entry)

DE Streptomyces avidinl1 sps protein.

KW Plant somatic tissue degeneration; plant essential factor; depletion;

KM viability; sps gene; plant development; plant morphology; flower;

KW fruit plant.

OS Streptomyces avidinl1.

PN WO200007427-A2.

PD 17-FEB-2000.

PF 30-JUL-1999; 99WO-IL00420.

XX 03-AUG-1998; 98IL-0125632.  
XX  
XX (AGRI-) AGRIC RES ORG.  
XX  
XX Kapulnik Y, Ginzberg I;  
XX  
XX WPI: 2000-195402/17.  
DR N-PSDB: 291073.  
XX  
XX Degeneration of somatic plant tissue by expression of a heterologous  
PT protein, useful for controlling plant development and morphology, such  
PT as decreasing the number of flowers present to increase the number of  
PT fruit -  
XX  
XX Examples: Page 84; 91pp; English.  
PS  
XX  
XX

CC The invention relates to a method of effecting degeneration of a somatic  
CC plant tissue by expressing a heterologous protein capable of binding a  
CC plant essential factor (PEF), in somatic plant tissue cells, where  
CC heterologous protein expression causes depletion of the PEF so the plant  
CC viability is maintained, while simultaneous degeneration of the somatic  
CC plant tissue is effected. This sequence represents the Streptomyces  
CC avidinl1 sps protein as an example of a heterologous protein introduced  
CC into the plants. The methods can provide for the selective and optionally  
CC reversible cell degeneration in somatic plant tissue. They can be used  
CC for artificially controlling plant development and morphology. They can  
CC be used e.g. to decrease the number of flowers in fruit producing plants  
CC so as to increase the number of fruits which reach maturity.  
XX  
SQ Sequence 186 AA:

Query Match 100.0%; Score 936; DB 21; Length 186;

Best Local Similarity 100.0%; Pred. No. 2.5e-76;

Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRIIVAAIAVSLTTVSITNASADPSKDSKAQVSAEAGITGTWYNQSGTFTVTAGAD 60  
DB 4 mklivvaalavslitvtasasadpskdskaqvsaaegitgtwngqstftvttagad 63  
OY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120  
DB 64 galgtgyesaavgnaesryvltgrysdapatdgsctalgtwvmaknnyrnhsattwsqgy 123  
OY 121 VGGAERINTQWLLTSGTTEANAMKSTLVGHDTFTYVKPSAASIDAARKAGVNNGNPLDA 180  
DB 124 vggaearrintqwilltsgtteanawkstlvghdtftkvkpsaasidaarkagvnngnplda 183  
OY 181 YQQ 183  
DB 181 yqq 183

RESULT 9

Y80513 Y80513 standard; Protein: 184 AA.

AC Y80513;

DT 06-JUN-2000 (first entry)

DE Streptomyces avidinl1 mst protein.

KW Plant somatic tissue degeneration; plant essential factor; depletion;

KM viability; mst gene; plant development; plant morphology; flower;

KW fruit plant.

OS Streptomyces avidinl1.

PN WO200007427-A2.

PD 17-FEB-2000.

XX 30-JUL-1999; 99WO-IL00420.  
 PF  
 XX  
 PR 03-AUG-1998; 98IL-0125632.  
 XX  
 PA (AGRI-) AGRIC RES ORG.  
 XX  
 PI Kapulnik Y, Ginzberg I;  
 XX  
 DR WPI: 2000-195402/17.  
 DR N-PSDB: Z91074.  
 XX  
 PT Degeneration of somatic plant tissue by expression of a heterologous  
 PT protein, useful for controlling plant development and morphology, such  
 PT as decreasing the number of flowers present to increase the number of  
 PT fruit -  
 XX  
 PS Examples: Page 85; 91pp; English.  
 XX  
 CC The invention relates to a method of effecting degeneration of a somatic  
 CC plant tissue by expressing a heterologous protein capable of binding a  
 CC plant essential factor (PEF), in somatic plant tissue cells, where  
 CC heterologous protein expression causes depletion of the PEF so the plant  
 CC viability is maintained, while simultaneous degeneration of the somatic  
 CC plant tissue is effected. This sequence represents the Streptomyces  
 CC avidin1 mat protein as an example of a heterologous protein introduced  
 CC into the plants. The methods can provide for the selective and optionally  
 CC reversible cell degeneration in somatic plant tissue. They can be used  
 CC for artificially controlling plant development and morphology. They can  
 CC be used e.g. to decrease the number of flowers in fruit producing plants  
 CC so as to increase the number of fruits which reach maturity.  
 XX  
 SQ Sequence 184 AA:

Query Match 99.5%; Score 931; DB 21; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 7e-76;  
 Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RKIVAAIAVSLTTVTSITTSASADPSKDSKAQYSAEAGITGTWYNQLGSTFTVTAGADG 61  
 |||||||  
 DB 3 rkivaaiaivsiltvsttsasadpskdskaqysaaegitgtwynqlgstftvtagadg 62  
 QY 62 ALTGTESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVAAMKNNYRNAHSATTMSGQYV 121  
 |||||||  
 DB 63 altgtesavгнаesryvltgrydsapadgsctalgwtvawknnyrnhsattmsgqyv 122  
 QY 122 GGAEARINTQWILTSGTTEANAMKSTLVGHDPFTTKVPSAASIDAAKKAGVNNGNPLDAY 181  
 |||||||  
 DB 123 ggaearintcqwiltsgtteanawkstlvghdfttkvpsaasidaakkagvnngnplday 182  
 QY 182 QQ 183  
 ||  
 DB 183 qq 184

RESULT 10  
 ID P70492 standard; protein; 182 AA.  
 XX  
 AC P70492;  
 XX  
 DT 06-MAR-1991 (first entry)  
 XX  
 DE Streptavidin sequence.  
 XX  
 KW Streptavidin; N-terminal; fusion gene; fusion protein;  
 XX  
 OS Streptomyces avidin1.  
 XX  
 PN W08705026-A.  
 XX  
 PD 27-AUG-1987.

XX 24-FEB-1987; 87WO-US00397.  
 PF  
 XX  
 PR 24-FEB-1986; 86US-0833324.  
 XX  
 PA (UYCO-) COLUMBIA UNIV N Y.  
 XX  
 PI (UYNX-) UNIV NEW YORK.  
 XX  
 PT Cantor CR, Axel R, Garana C;  
 XX  
 DR WPI: 1987-250198/35.  
 DR N-PSDB: N70810.  
 XX  
 PT DNA encoding streptavidin - obtd. by restriction endo-nuclease  
 PT digestion of chromosomal DNA of Streptomyces avidin1  
 XX  
 PS Disclosure; Figure 3; 54pp; English.  
 XX  
 CC The sequence is that of streptavidin from S.avidin1. Streptavidin  
 CC may be expressed from a fusion gene comprising its coding gene and DNA  
 CC encoding a target protein of interest, where the streptavidin has  
 CC binding sites for biotin or deriv. The streptavidin has 4 binding  
 CC sites free for biotin, and is produced free of biotin  
 CC contamination. Improved streptavidins may also be produced by  
 CC site-directed mutagenesis. The fused gene may be used to produce  
 CC labeled, chemically-modified proteins in vivo, and to isolate  
 CC proteins when only the sequence of the gene is known.  
 XX  
 SQ Sequence 182 AA:

Query Match 99.1%; Score 928; DB 8; Length 182;  
 Best Local Similarity 99.5%; Pred. No. 1.3e-75;  
 Matches 181; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RKIVAAIAVSLTTVTSITTSASADPSKDSKAQYSAEAGITGTWYNQLGSTFTVTAGADG 61  
 |||||||  
 DB 1 rkivaaiaivsiltvsttsasadpskdskaqysaaegitgtwynqlgstftvtagadg 60  
 QY 62 ALTGTESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVAAMKNNYRNAHSATTMSGQYV 121  
 |||||||  
 DB 61 altgtesavгнаesryvltgrydsapadgsctalgwtvawknnyrnhsattmsgyv 120  
 QY 122 GGAEARINTQWILTSGTTEANAMKSTLVGHDPFTTKVPSAASIDAAKKAGVNNGNPLDAY 181  
 |||||||  
 DB 121 ggaearintcqwiltsgtteanawkstlvghdfttkvpsaasidaakkagvnngnplday 180  
 QY 182 QQ 183  
 ||  
 DB 181 qq 182

RESULT 11  
 ID Y84020 standard; protein; 183 AA.  
 XX  
 AC Y84020;  
 XX  
 DT 03-JUL-2000 (first entry)  
 XX  
 DE Amino acid sequence of the wildtype streptavidin monomer.  
 XX  
 KW Streptavidin; monomer; biotin binding domain; functional domain;  
 KW biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;  
 KW substrate; vascular device; prosthesis.  
 XX  
 OS Streptomyces avidin1.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..24  
 FT /note= "propeptide"  
 XX  
 PN W0200011152-A1.

XX 02-MAR-2000.  
 XX  
 PD  
 XX  
 PF 25-AUG-1999; 99MO-US19481.  
 XX  
 PR 25-AUG-1998; 98US-0097816.  
 XX  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Strayton PS, McDevitt TC, Nelson KE;  
 XX WPI: 2000-224689/19.  
 DR  
 XX  
 PT Mutant streptavidin molecule, useful as an adaptor and a coating for  
 PT substrates such as vascular devices or prostheses, comprises a biotin  
 PT binding domain and a secondary functional domain -  
 PS Disclosure; Page 35-36; 43pp; English.  
 XX  
 CC The present sequence represents a wildtype streptavidin monomer. The  
 CC specification describes streptavidin molecules comprising a biotin  
 CC binding domain and a secondary functional domain. The molecules are  
 CC adaptors with inherent effector function. They can therefore, bind to  
 CC biotin and also have another function, for e.g. binding to a cell  
 CC through the secondary functional domain comprising a cell adhesion  
 CC peptide. The streptavidin molecules are useful as adaptors to bring,  
 CC via a streptavidin/biotin interaction, the secondary functional domain  
 CC into proximity with a cell or molecule to be affected and as a coating  
 CC for substrates such as vascular devices or prostheses. Therefore, any  
 CC compound of interest, such as a nucleic acid, protein, peptide, organic  
 CC compound, inorganic compound, polysaccharide or a combination, can be  
 CC targeted, delivered or immobilized using them. The molecules, when  
 CC comprising an antibody for its secondary functional domain, are also  
 CC useful in diagnostic applications for detecting analytes.  
 XX  
 XX Sequence 183 AA:  
 SQ  
 Query Match 99.1%; Score 928; DB 21; Length 183;  
 Best Local Similarity 99.5%; Pred. No. 1.3e-75;  
 Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MRKIVAAINSLTFTVITTSASADPSKDSKAQYSAEAGITGWNQOLGFTFVTAGAD 60  
 Db 1 mrklvvaalavslttvlttaasasdpkdkagysaaegltgwnqlgscflvtagsad 60  
 QY 61 GALTGYESAVGNAESRYVLTGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120  
 Db 61 galctgyesavgnaesryvltgrydsapctdgsctalgwtvawknynrnhsattwsqgy 120  
 QY 121 VGGAEARINTOMLITSGTTEANAMKSTLVGHDFRTKYPKPSAASIDAACKKAGVNNGNPLDA 180  
 Db 121 vggaearinltqwtltsgteanawksclvhdfctkvkpsaasidaackkagvnnngnplda 180  
 QY 181 VQO 183  
 Db 181 vqg 183  
 RESULT 12  
 W59217  
 ID W59217 standard; Protein: 183 AA.  
 AC W59217;  
 XX  
 DT 27-AUG-1998 (first entry)  
 XX  
 XX S. avidinli streptavidin mutant protein #1.  
 XX  
 KW Streptavidin: ligand; binding affinity; mutant; isolation;  
 KM purification; recover; immobilise.  
 XX  
 OS Streptomyces avidinli.

OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 68..71  
 FT /label= VTAR68ESAV  
 FT /note= "Wild type ESAV is replaced by VTAR. Numbering  
 FT is from the start of the mature protein"  
 XX  
 XX EP835934-A2.  
 XX  
 PD 15-APR-1998.  
 XX  
 PF 09-OCT-1997; 97EP-0117504.  
 XX  
 PR 10-OCT-1996; 96DE-1041876.  
 XX  
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.  
 PI Skerra A, Voss S;  
 XX WPI: 1998-218868/20.  
 DR N-PSDB; V34715.  
 XX  
 XX  
 PT Streptavidin mutants with higher binding affinity for peptide  
 PT ligands - have mutation in amino acid region 44-53, used to isolate,  
 PT purify or determine fusion proteins including these ligands  
 PS Disclosure; Page -: 21pp; German.  
 XX  
 XX This sequence represents a mutant streptavidin protein isolated from  
 CC Streptomyces avidinli where the residues ESAV at position 44-47 of  
 CC the mature wild type sequence are replaced by VTAR. This sequence is  
 CC used to produce mutants which are used in a method to assay the binding  
 CC affinity of streptavidin mutants. These mutants have a mutation within  
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher  
 CC binding affinity than the wild-type for peptide ligands that include  
 CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y  
 CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant  
 CC streptavidin mutants can be used to isolate, purify and determine  
 CC proteins or to determine/recover substances that contain  
 CC streptavidin-binding groups. Such compounds may also be used to  
 CC immobilise fusions on microtitre plates, microbeads or sensor chips.  
 CC NOTE: this sequence does not appear in the specification but has  
 CC been constructed from the wild-type streptavidin sequence represented  
 CC in V34714.  
 XX  
 SQ Sequence 183 AA:  
 Query Match 98.2%; Score 919; DB 19; Length 183;  
 Best Local Similarity 98.4%; Pred. No. 8.2e-75;  
 Matches 180; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MRKIVAAINSLTFTVITTSASADPSKDSKAQYSAEAGITGWNQOLGFTFVTAGAD 60  
 Db 1 mrklvvaalavslttvlttaasasdpkdkagysaaegltgwnqlgscflvtagsad 60  
 QY 61 GALTGYESAVGNAESRYVLTGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120  
 Db 61 galctgyvltaragneesryvltgrydsapctdgsctalgwtvawknynrnhsattwsqgy 120  
 QY 121 VGGAEARINTOMLITSGTTEANAMKSTLVGHDFRTKYPKPSAASIDAACKKAGVNNGNPLDA 180  
 Db 121 vggaearinltqwtltsgteanawksclvhdfctkvkpsaasidaackkagvnnngnplda 180  
 QY 181 VQO 183  
 Db 181 vqg 183  
 RESULT 13  
 W59218  
 ID W59218 standard; Protein: 183 AA.

```
XX AC W59218;
XX DT 27-AUG-1998 (first entry)
XX DE S. avidinii streptavidin mutant protein #2.
XX KM Streptavidin; ligand; binding affinity; mutant; isolation;
XX KM purification; recover; immobilise.
XX OS Streptomyces avidinii.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 68..71
XX FT /label= IGAR68ESAV
XX FT /note= "Wild type ESAV is replaced by IGAR. Numbering
FT is from the start of the mature protein"
XX PN EP835934-A2.
XX PD 15-APR-1998.
XX PF 09-OCT-1997; 97EP-0117504.
XX PR 10-OCT-1996; 96DE-1041876.
XX PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX PI Skerra A, Voss S;
XX DR WPI; 1998-218868/20.
XX DR N-PSDB; V34716.
XX PT Streptavidin mutants with higher binding affinity for peptide
XX PT ligands - have mutation in amino acid region 44-53, used to isolate,
XX PT purify or determine fusion proteins including these ligands
XX PS Disclosure; Page -: 21pp; German.
XX CC This sequence represents a mutant streptavidin protein isolated from
XX CC Streptomyces avidinii where the residues ESAV at position 44-47 of
XX CC the mature wild type sequence are replaced by IGAR. This sequence is
XX CC used to produce mutants which are used in a method to assay the binding
XX CC affinity of streptavidin mutants. These mutants have a mutation within
XX CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX CC binding affinity than the wild-type for peptide ligands that include
XX CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y
XX CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant
XX CC streptavidin mutants can be used to isolate, purify and determine
XX CC proteins or to determine/recover substances that contain
XX CC streptavidin-binding groups. Such compounds may also be used to
XX CC immobilise fusions on microtitre plates, microbeads or sensor chips.
XX CC NOTE: This sequence does not appear in the specification but has
XX CC been constructed from the wild-type streptavidin sequence represented
XX CC in V34714.
XX SQ Sequence 183 AA;

Query Match 98.0%; Score 917; DB 19; Length 183;
Best Local Similarity 98.4%; Pred. No. 1.2e-74;
Matches 180; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Db 121 vggaeairintqwltsgtteanmkstlvghdftvkvkpsaasidaakagvnmnplda 180
QY 181 VQQ 183
Db 181 vqg 183

RESULT 14
Y84021
ID Y84021 standard; protein; 186 AA.
XX AC
XX DT Y84021;
XX DE 03-JUL-2000 (first entry)
XX DE Amino acid sequence of a fibrinonectin/streptavidin fusion.
XX KM Streptavidin; monomer; biotin binding domain; functional domain;
XX KM biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;
XX KM substrate; vascular device; prosthesis; fibrinonectin.
XX OS Synthetic.
XX OS Streptomyces avidinii.
XX FH Key Location/Qualifiers
XX FT Peptide 1..24
XX FT /note= "propeptide"
XX PN W020001152-A1.
XX PD 02-MAR-2000.
XX PF 25-AUG-1999; 99WO-US19481.
XX PR 25-AUG-1998; 98US-0097816.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Stayton PS, McDevitt TC, Nelson KE;
XX DR WPI; 2000-224689/19.
XX PT Mutant streptavidin molecule, useful as an adaptor and a coating for
XX PT substrates such as vascular devices or prostheses, comprises a biotin
XX PT binding domain and a secondary functional domain -
XX PS Example 4; Page 36-37; 43pp; English.
XX CC The present sequence represents a fusion of fibrinonectin and
XX CC streptavidin. It is a streptavidin molecule of the invention. The
XX CC specification describes streptavidin molecules comprising a biotin
XX CC binding domain and a secondary functional domain. The molecules are
XX CC adaptors with inherent effector function. They can therefore bind to
XX CC biotin and also have another function, for e.g. binding to a cell
XX CC through the secondary functional domain comprising a cell adhesion
XX CC peptide. The streptavidin molecules are useful as adaptors to bring,
XX CC via a streptavidin/biotin interaction, the secondary functional domain
XX CC into proximity with a cell or molecule to be affected and as a coating
XX CC for substrates such as vascular devices or prostheses. Therefore, any
XX CC compound of interest, such as a nucleic acid, protein, peptide, organic
XX CC compound, inorganic compound, polysaccharide or a combination, can be
XX CC targeted, delivered or immobilized using them. The molecules, when
XX CC comprising an antibody for its secondary functional domain, are also
XX CC useful in diagnostic applications for detecting analytes.
XX SQ Sequence 186 AA;

Query Match 96.3%; Score 901.5; DB 21; Length 186;
Best Local Similarity 96.2%; Pred. No. 3e-73;
Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
```

QY 1 MKRIVVAALAVSLTIVTSITASADSPSKDSKAQVSAEAAGITGTWYNQLGSTFIYTAGAD 60  
 DB 1 mrklvvaalavslitvslitasadpskdskaqvsaaagltglwvnyqlgstflivtagad 60  
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPA---TDGSGTALGWTAVAMKNRYRNAHSATTWS 117  
 DB 61 galtgtyesavgnaesryvltgrysdapgsgtalgwtvawkmnyrnhsattws 120  
 QY 118 GOYVGAEARINTQWLLTSGTTEANAMKSTLVGHDFTTKVPSAASIDAKKAGVNNGNP 177  
 DB 121 ggyvgsaearintqwalltsgtteanawkscllvhdfttkvpsaasidaakkagvnngrp 180  
 QY 178 LDAVQQ 183  
 DB 181 ldavqg 186

RESULT 15  
 Y84022  
 Y84022 standard; protein: 186 AA.  
 XX  
 AC Y84022;  
 XX  
 DT 03-JUL-2000 (first entry)  
 XX  
 DE Amino acid sequence of a osteopontin/streptavidin fusion.  
 XX  
 KW Streptavidin; monomer; biotin binding domain; functional domain;  
 KW biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;  
 KW substrate; vascular device; prosthesis; osteopontin.  
 XX  
 OS Synthetic.  
 OS Streptomyces avidin11.  
 FH Key Location/Qualifiers  
 FT 1..24  
 FT Peptide /note= "propeptide"  
 XX  
 PN MO200011152-A1.  
 XX  
 PD 02-MAR-2000.  
 XX  
 PF 25-AUG-1999; 99WO-US19481.  
 XX  
 PR 25-AUG-1998; 98US-0097816.  
 XX  
 PA (UNITV ) UNIV WASHINGTON.  
 PI Staylor PS, McDevitt TC, Nelson KE;  
 XX WPI; 2000-224689/19.  
 DR  
 XX  
 PT Mutant streptavidin molecule, useful as an adaptor and a coating for  
 PT substrates such as vascular devices or prostheses, comprises a biotin  
 PT binding domain and a secondary functional domain -  
 XX  
 PS Disclosure: Page 38; 43pp: English.  
 XX  
 CC The present sequence represents a fusion of osteopontin and  
 CC streptavidin. It is a streptavidin molecule of the invention. The  
 CC specification describes streptavidin molecules comprising a biotin  
 CC binding domain and a secondary functional domain. The molecules are  
 CC adaptors with inherent effector function. They can therefore, bind to  
 CC biotin and also have another function, for e.g. binding to a cell  
 CC through the secondary functional domain comprising a cell adhesion  
 CC peptide. The streptavidin molecules are useful as adaptors to bring,  
 CC via a streptavidin/biotin interaction, the secondary functional domain  
 CC into proximity with a cell or molecule to be affected and as a coating  
 CC for substrates such as vascular devices or prostheses. Therefore, any  
 CC compound of interest, such as a nucleic acid, protein, peptide, organic  
 CC compound, inorganic compound, polysaccharide or a combination, can be  
 CC targeted, delivered or immobilized using them. The molecules, when  
 CC comprising an antibody for its secondary functional domain, are also

CC useful in diagnostic applications for detecting analytes.  
 XX  
 SQ Sequence 186 AA;

Query Match 96.1%; Score 899.5; DB 21; Length 186;  
 Best Local Similarity 96.2%; Pred. No. 4, 6e-73;  
 Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MKRIVVAALAVSLTIVTSITASADSPSKDSKAQVSAEAAGITGTWYNQLGSTFIYTAGAD 60  
 DB 1 mrklvvaalavslitvslitasadpskdskaqvsaaagltglwvnyqlgstflivtagad 60  
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPA---TDGSGTALGWTAVAMKNRYRNAHSATTWS 117  
 DB 61 galtgtyesavgnaesryvltgrysdapgsgtalgwtvawkmnyrnhsattws 120  
 QY 118 GOYVGAEARINTQWLLTSGTTEANAMKSTLVGHDFTTKVPSAASIDAKKAGVNNGNP 177  
 DB 121 ggyvgsaearintqwalltsgtteanawkscllvhdfttkvpsaasidaakkagvnngrp 180  
 QY 178 LDAVQQ 183  
 DB 181 ldavqg 186

Search completed: April 6, 2001, 00:15:59  
 Job time: 15499 sec







```
|||||
250 CGAGTCGGCCCTCGGCAACGCCGAGAGCCGCTACGCTCGACCGGTCGT 299
84 YRASPserAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
300 ACGACAGGCCCCGGCCACCGACGCGACGCGCCCTCGGTGGAGC 349
101 ValAlaTrpLysAsnAsnTyraArgAsnAlaHisSerAlaThrThrTrpSe 117
350 GTGGCTGTGAAGAAATTAATACCGCAACCCCACTCCGGACACGCTGAG 399
117 rGlyGlnTyraValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
400 CGGCGACGACGTCGGCGCGCGCGAGCGAGATCAACACCCAGTGGCTGC 449
134 eutThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
450 TGACCTCCGGCACCAACGCGCAACGCTGGAAGTCCACGCTGGTCGCG 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
500 CACGACACCTTCACCAAGGTGAAGCCGTCGCGCTCCATCGACGCGCG 549
167 aLysLysAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
550 GAAGAAGGCGCGCGCTCAACAACGCAACCCGCTCGACGCGCTCGACGAG 598
```

seq\_name: gb\_pat1:A93649

seq\_documentation\_block:

LOCUS A93649 638 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 1 from Patent EP0799890.  
ACCESSION A93649  
VERSION A93649.1 GI:6741838

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE  
1 (bases 1 to 638)  
AUTHORS Mueller,R.D. and Deger,A.D.  
TITLE Recombinant inactive core streptavidin mutants  
JOURNAL Patent: EP 0799890-A 08-OCT-1997;  
BOEHRINGER MANNHEIM GMBH (DE)

FEATURES  
source Location/Qualifiers  
1..638  
sig\_peptide /organism="unidentified"  
CDS /db\_xref="taxon:32644"  
50..121  
50..601  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAB69587.1"  
/db\_xref="GI:6741839"  
/translation="MKIVVAIAVSLTTSITASADPSKDSKAQVSAAPAGITGT  
WYNQGSTPIYTAGADALGTYESAVGNAESRYLTGRYDAPAPDGSGLAGWTVA  
WNNRNRAHSATTTWSGOYVGAEARINTQMLTSGTEANAMKSTLVGHDTFTKVKPS  
ASIDAARKAGVNGNPLDAVQO"

mat\_peptide 122..598  
BASE COUNT 115 a 244 c 193 g 86 t  
ORIGIN

alignment\_scores:

Quality: 936.00 Length: 183  
Ratio: 5.115 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-589-870-2 x A93649 ..

Align seg 1/1 to: A93649 from: 1 to: 638

1 MejArGlySileValAlaAlaIleAlaValSerLeuThrThrValSe 17

```
|||||
50 ATGCGCAAGATCGTCGTCGACCATGCGCTTCCTCCGACCAAGGTCCTC 99
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
100 GATTACGGCCAGCCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAAG 149
34 aLSerAlaAlaGluAlaGlyIleThrGlyThrTrpTyraSngLleuGly 50
150 TCTGGCCCGCGAGCGCGCATCAACGGACCTGATACACCACTCGGC 199
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
200 TCGACCTTCATCGTGACCGCGCGCGCGACGCGCCCTCGACGGAACTA 249
67 rGluSerAlaValGlyLysAlaGluSerArTyraValLeuThrGlyArgT 84
250 CGAGTCGGCCGTCGGCAACGCGAGCGCGCTACGTCCTGACCGGTGTT 299
84 YRASPserAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
300 ACGACAGGCCCCGGCCACGACGCGACGCGCGCTCGGTGGAGC 349
101 ValAlaTrpLysAsnAsnTyraArgAsnAlaHisSerAlaThrThrTrpSe 117
350 GTGGCTGTGAAGAAATTAATACCGCAACCCCACTCCGGACACGCTGAG 399
117 rGlyGlnTyraValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
400 CGGCGACGACGTCGGCGCGCGCGAGCGAGATCAACACCCAGTGGCTGC 449
134 eutThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
450 TGACCTCCGGCACCAACGCGCAACGCTGGAAGTCCACGCTGGTCGCG 499
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
500 CACGACACCTTCACCAAGGTGAAGCCGTCGCGCTCCATCGACGCGCGC 549
167 aLysLysAlaGlyValAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
550 GAAGAAGGCGCGCGCTCAACAACGCAACCCGCTCGACGCGCTCGACGAG 598
```

seq\_name: gb\_pat1:101349

seq\_documentation\_block:

LOCUS 101349 638 bp ss-DNA PAT 21-MAY-1993  
DEFINITION Sequence 2 from Patent US 4839293.  
ACCESSION 101349  
VERSION 101349.1 GI:270135

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE  
1 (bases 1 to 638)  
AUTHORS Cantor,C.R., Axel,R. and Argarana,C.  
TITLE DNA encoding streptavidin, streptavidin produced therefrom, fused  
polypeptides which include amino acid sequences present in  
streptavidin and uses thereof  
JOURNAL Patent: US 4839293-A 2 13-JUN-1989;  
The Trustees of Columbia University in the City of New York;  
New York, NY

FEATURES  
source Location/Qualifiers  
1..638  
BASE COUNT 114 a 244 c 194 g 86 t  
ORIGIN

alignment\_scores:

Quality: 933.00 Length: 183  
Ratio: 5.098 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.454

## Alignment block:

US-09-589-870-2 x I01349 ..

Align seg 1/1 to: I01349 from: 1 to: 638

```

1 MetArgLysIleValIAlaIleAlaIleAlaValSerLeuThrThrValSe 17
  |||||||
50 ATGGCGCAAGATCGTCGTTGCAGCCATCGCCCTTCCCTGACGACGCTCTC 99
  |||||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
  |||||||
100 GATTACGGCCAGCCGCTTCGCGACACCCTCCAGAGACTCGAGGGCCAGG 149
  |||||||
34 aIleSerAlaIleAlaGlnAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
  |||||||
150 TCTCGCGCCGCGAGCGCGCATACCGGCACCTGCTACACACCTCTGGC 199
  |||||||
51 SerThrPheIleValIThrAlaGlyAlaIleAspGlyAlaLeuThrGlyThrTy 67
  |||||||
200 TCGACCTTCATCGTCGACCGCGCGCCGCGACGCGCCCTGACCGGAACTTA 249
  |||||||
67 rGluSerAlaValAlaGlyAlaSerArgTyrValLeuThrGlyArgT 84
  |||||||
250 CGAGTCGGCGCGTCGCGCAACGCCGAGAGCGCTTACGCTCGACCGGTGCTT 299
  |||||||
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
  |||||||
300 ACGACAGCGCGCCGCGCCACGCGACGCGCGACCGCCCTCGGTGGAGC 349
  |||||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
  |||||||
350 GTGGCCTCGAAGAACTAACTACCGCAACGCCCACTCCGCGACGCTGGAG 399
  |||||||
117 rGlyGlnTyrValAlaGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuL 134
  |||||||
400 CGGCAGTACGTCGCGCGCGCGAGGAGATCAACACCCAGTGGCTGTC 449
  |||||||
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
  |||||||
450 TGACCTCGCGCACCAACGAGGCAACGCTGGAAGTCCACGCTGGTCTGGC 499
  |||||||
151 HisAspThrPheThrLysValLysProSerAlaIleAspSerIleAspAlaI 167
  |||||||
500 CACGACACCTTCACCAAGGTGAAGCCGTCGCCGCTCATCGACGCGGC 549
  |||||||
167 alyLysAlaGlyAlaValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
  |||||||
550 GAAGAAGCGCGCGCTCAACAACGGAACCGCGCTCGACGCGGTTCAGCAG 598
  |||||||

```

seq\_name: gb\_ba3:S78777

seq\_documentation\_block:

LOCUS S78777 625 bp DNA BCT 30-OCT-1995  
 DEFINITION streptavidin v1 [Streptomyces violaceus, Tu 2460, Genomic, 625 nt].  
 ACCESSION S78777  
 VERSION S78777.1 GI:1042193  
 KEYWORDS  
 SOURCE Streptomyces violaceus Tu 2460.  
 ORGANISM Streptomyces violaceus  
 Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
 Actinomycetales: Streptomycineae: Streptomycetaceae: Streptomyces.

REFERENCE 1 (bases 1 to 625)

AUTHORS Bayer, E.A., Kulik, T., Adar, R. and Wilchek, M.

TITLE Close similarity among streptavidin-like, biotin-binding proteins from Streptomyces

JOURNAL Biochim. Biophys. Acta 1263 (1), 60-66 (1995)

MEDLINE Genbank staff at the National Library of Medicine created this

REMARK entry [NCBI g1bbsq 169185] from the original journal article.

FEATURES This sequence comes from Fig. 4.

source Location/Qualifiers

1..625

```

/organism="Streptomyces violaceus"
/db_xref="taxon:1936"
50..601
/gene="streptavidin v1, Sa v1"
50..601
/gene="streptavidin v1, Sa v1"
/note="This sequence comes from Fig. 5; Sa v1"
/codon_start=1
/product="streptavidin v1"
/protein_id="AAB3501.1"
/db_xref="GI:1042194"
/translation="MKRIYVAALAVSLTFTVSTIASADPSKSKAQAQVAEAGITGT
WYNQGSFTVTRAGADALTGTEASAVGNESRYVLTGRYDSAPATDGSSTALGMTVA
WKNRYNAHSATYMGQYGVGTGARINTQMLTLSTGTANAMKSTLVGHDTFTKVRPS
ASIDAKKAGVNNKPNLDAVD0"
BASE COUNT 115 a 240 c 186 g 84 t
ORIGIN

```

```

alignment_scores:
  Quality: 932.00 Length: 183
  Ratio: 5.093 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.454

```

## alignment block:

US-09-589-870-2 x S78777 ..

Align seg 1/1 to: S78777 from: 1 to: 625

```

1 MetArgLysIleValIAlaIleAlaIleAlaValSerLeuThrThrValSe 17
  |||||||
50 ATGGCGCAAGATCGTCGTTGCAGCCATCGCCCTTCCCTGACGACGCTCTC 99
  |||||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
  |||||||
100 GATTACGGCCAGCCGCTTCGCGACACCCTCCAGAGACTCGAGGGCCAGG 149
  |||||||
100 GATTACCGCCAGCCGCTTCGCGACACCCTCCAGAGACTCGAGGGCCAGG 149
  |||||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
  |||||||
100 GATTACCGCCAGCCGCTTCGCGACACCCTCCAGAGACTCGAGGGCCAGG 149
  |||||||
34 aIleSerAlaIleAlaGlnAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
  |||||||
150 TCTCGCGCCGCGAGCGCGCATACCGGCACCTGCTTACACACGCTCGGC 199
  |||||||
51 SerThrPheIleValIThrAlaGlyAlaIleAspGlyAlaLeuThrGlyThrTy 67
  |||||||
200 TCGACCTTCATCGTCGACCGCGCGCCGCGACGCGCCCTGACCGGAACTTA 249
  |||||||
67 rGluSerAlaValAlaGlyAlaSerArgTyrValLeuThrGlyArgT 84
  |||||||
250 CGAGTCGGCGCGTCGCGCAACGCCGAGAGCGCGCTTACGCTCGACCGGTGCTT 299
  |||||||
84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
  |||||||
300 ACGACAGCGCGCCGCGCCACGCGACGCGCGACCGCCCTCGGTGGAGC 349
  |||||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
  |||||||
350 GTGGCCTCGAAGAACTAACTACCGCAACGCCCACTCCGCGACGCTGGAG 399
  |||||||
117 rGlyGlnTyrValAlaGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuL 134
  |||||||
400 CGGCAGTACGTCGCGCGCGCGAGGAGATCAACACCCAGTGGCTGTC 449
  |||||||
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
  |||||||
450 TGACCTCGCGCACCAACGAGGCAACGCTGGAAGTCCACGCTGGTCTGGC 499
  |||||||
151 HisAspThrPheThrLysValLysProSerAlaIleAspSerIleAspAlaI 167
  |||||||
500 CACGACACCTTCACCAAGGTGAAGCCGTCGCCGCTCATCGACGCGGC 549
  |||||||
167 alyLysAlaGlyAlaValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
  |||||||
550 GAAGAAGCGCGCGCTCAACAACGGAACCGCGCTCGACGCGGTTCAGCAG 598
  |||||||

```

```
seq_name: gb_ba3:S78782
seq_documentation block:
LOCUS       S78782             625 bp    DNA             BCT       30-OCT-1995
DEFINITION   streptavidin v2 [Streptomyces violaceus, Tu 2605, Genom1c, 625 nt].
ACCESSION    S78782
VERSION      S78782.1 GI:1042195
KEYWORDS
SOURCE       Streptomyces violaceus Tu 2605.
ORGANISM     Streptomyces violaceus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.
REFERENCE    1 (bases 1 to 625)
AUTHORS      Bayer, E.A., Kulik, T., Adar, R. and Wilchek, M.
TITLE        Close similarity among streptavidin-like, biotin-binding proteins
            from Streptomyces
JOURNAL      Biochim. Biophys. Acta 1263 (1), 60-66 (1995)
MEDLINE      93359204
REMARK       Genbank staff at the National Library of Medicine created this
            entry [NCBI g1bseq 169186] from the original journal article.
            This sequence comes from Fig. 4.
FEATURES
     source
         location/Qualifiers
             1..625
             /organism="Streptomyces violaceus"
             /db_xref="taxon:1936"
             /gene="streptavidin v2, Sa v2"
             /note="This sequence comes from Fig. 5; Sa v2"
             /codon_start=1
             /product="streptavidin v2"
             /protein_id="AAB35016.1"
             /db_xref="GI:1042196"
             /translation="MKYIVAAIAVSLTVGVTASASAPDSKSKQAQVAEAGTGT
             WYNQGSTVTIVANADGSLTGTFESAVAESRYVLTGYDAPADGSGTALGWTV
             WNNIRNASHATWISGOYVAGSEARINTQMLTSGTTAANAMKSTLVGHDFTFKYPS
             AASIDAARKAGVNNGNPLDAYOQ"
BASE COUNT   117 a      244 c      184 g      80 t
ORIGIN
alignment_scores:
    Quality:      897.00      Length:      183
    Ratio:        4.929      Gaps:      0
    Percent Similarity: 99.454      Percent Identity: 95.082
alignment block:
US-09-589-870-2 x S78782 ..
Align seg 1/1 to: S78782 from: 1 to: 625
1 MetArgLysIleValAlaAlaIleAlaValSerIleuThrValSe 17
|||||
50 ATGCGCAGAGATCGTGTGACGACCATTCCTTCGTGACACCGTCGG 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGly 34
|||||
100 TATTAAGCGCAGCGCCCTCCGGGACCCGTCGAAGACTCCACAGGCCAGG 149
|||||
34 aISerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50
|||||
150 CCGCGCTTCGCGGAGCGCGCATCCGCGACCTGTCACACACACTCGGT 199
|||||
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyr 67
|||||
200 TCGACCTTCAATCGACCGGCAAGCGCGACGACCTCAACCGGACACTA 249
|||||
67 rGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValIleuThrGlyArgT 84
|||||
250 CGAGTCGGCGGTGCGCAACCGCGAGACCGCTACGCTCAACCGGTGTT 299
|||||
- 84 YrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
```

```
|||||
300 ACGACAGCGCCCGCGCCACTGACGCGACGCGACCGCCCTGCGTGGACG 349
|||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSe 117
|||||
350 GTGGCGTGAAGAACACTACCGCAACGCCCACTCCGCCACACGCTGAG 399
|||||
117 rGlyGlnTrpValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuL 134
|||||
400 CGCGCAATACGTCCGCGGCTCCGAGCGCCGATCAACACCCAGTGGCTGC 449
|||||
134 eufHrSerGlyThrThrGlnAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCGGCGACACACCGCGCGCAACGCTGTGAAGTCCACGCTGTGCGC 499
|||||
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaI 167
|||||
500 CACGACACCTTCACCAAGGTGAAGCGCGTCCGCGGTCATCGACGCCGCG 549
|||||
167 aLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 CAAGAGGCGCGCGCTCAACAGCGCAACCCGCTCGACGCGTCCAGCAG 598
|||||
seq_name: gb_pat1:A39565
seq_documentation block:
LOCUS       A39565             1356 bp    DNA             PAT       05-MAR-1997
DEFINITION   Sequence 4 from Patent WO9415644.
ACCESSION    A39565
VERSION      A39565.1 GI:2295847
KEYWORDS
SOURCE       unidentified.
ORGANISM     unidentified.
REFERENCE    1 (bases 1 to 1356)
AUTHORS      Epenetos, A.A., Spooner, R.A. and Deonarain, M.
TITLE        COMPOUNDS FOR TARGETING
JOURNAL      Patent: WO 9415644-A 4 21-JUL-1994;
            IMP CANCER RES TECH (GB)
            Other publication GB 2289679 951129.
FEATURES
     source
         location/Qualifiers
             1..1356
             /organism="unidentified"
             /db_xref="taxon:32644"
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="CAA02449.1"
             /db_xref="GI:2295848"
             /translation="MKYLLPTAAAGLLILAAOPAMAQYOLOQPGAELYKPGASVYLSG
             KASGYPTSYMMHWKQRPGRLENGRILDRSGTKYNEKFKSKATLTVDKPSSTAY
             MQLSLTSEDSKAVYICARVDYGSSTFDYWGQTTVYSSGGGSGGGSGGGGQAV
             VTQESALTTSPGETVTLTCRSSTGAVTTSNANWVQKPDHLFTGLIGTNNRAGVP
             ARFSGSLIGDKAKALITGAQTEDEALIVCALMYSNHWVFGGTKLTIVGLEPAAPAP
             DPSKDSKAQVSAEAGITGTWYNGCSFETVYAGADALTCGYESAAGVNASRYVLTG
             RYDSPAIDGSGSTALGWTAVMKNNRNHNSATFWSGOYVGAEEARINTQMLTSGTTE
             ANAMKSTLVGHDFTFKYKPSAASIDAARKAGVNNGNPLDAYOQ"
BASE COUNT   332 a      357 c      361 g      306 t
ORIGIN
alignment_scores:
    Quality:      846.00      Length:      164
    Ratio:        5.190      Gaps:      0
    Percent Similarity: 99.390      Percent Identity: 98.780
alignment block:
US-09-589-870-2 x A39565 ..
Align seg 1/1 to: A39565 from: 1 to: 1356
20 AlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAl 36
|||||
```

```
853 GCTGCCGACCTGCAGACCCGTCACAGACTCCAAAGCTCAGTTTCTGC 902
36 AAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGlySerThr 53
903 AGCCGAAGCTGTATCACTGCGACCTGTATTAACCACTGGGCTGCACT 952
53 heileValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGlySer 69
953 TCATTGTGACCGCTGCTGCGACGAGACTCTGACTGGCAGCTACGATCT 1002
70 AlaValGlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSe 86
1003 GCGGTTGTAAACGAGAAATCCCGCTACGTAAGTCTGAGCCCTTATGACTC 1052
86 rAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPrpValAlaT 103
1053 TGCACCTGCCACCGATGCTGCTGACCGCTGCGCTGCGACTGCTGCTT 1102
103 rPlyAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSerGlyGln 119
1103 GGAAGAAACAACTATCTATATCCGACAGCCGCACTACGTGCTGCCAA 1152
120 TyrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSe 136
1153 TACGTTGGCGGCTGCTGAGGCTCGTATCAACACTCAGTGGCTGTTAACTC 1202
136 rGlyThrThrGluAlaAsnAlaTrpPlySerThrLeuValGlyHisAspT 153
1203 CGGCACTACCGAAGCAATGCATGAAATCGACACTAGTAGTCAATGACA 1252
153 hrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLys 169
1253 CCTTACCAGAACTTAAGCCTTCTGCTGCTGATGATGCTGCCAAGAAA 1302
170 AlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
1303 GCAGGCGGTAAACAGCGTAACCTCTAGAGCTGTTCAGCAA 1344
seq_name: gb_pat1:A93152
seq_documentation_block:
LOCUS A93152 1356 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 4 from Patent EP0815872.
ACCESSION A93152
VERSION A93152.1 GI:6741540
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos,A.A. and Deonarain,M.
TITLE Compounds for targeting
JOURNAL Patent: EP 0815872-A 07-JAN-1998;
IMP CANCER RES TECH (GB)
FEATURES
source Location/Qualifiers
1..1356
/organism="unidentified"
/db_xref="taxon:32644"
40..1347
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAB69540.1"
/db_xref="GI:6741541"
/translacion="MKYLLPTAAAGLLLLAOPAMAQVQLQPGAEIVKPGASVKLS
KASGYFTSYMMHWYKORPGRGLEWIGRIDPNSGRTYVSSGGSGSGGSGQAV
NOLSLTSEDSAVYYCARYDYGSYEDWGGTFTYSSGGSGSGGSGGSGQAV
VHQSALTSPEETVITCRSSGTAATSTNANMVEKPDHLEFTGLIGTNNRPPVP
AARPSGLIDKRAALITTCQTEDRAITFCALYKSNHWYGGGRTLVLDLEAPAAPA
DPSKDSKAOVSAAEAGITGTWVNDUSTFIVTAGADGALIGTESAVNAESRYVLTG
RVDSPATGSGTALGTVAMKNRNNAHSATTWSSQYVGAEARINTWQLTSLTSTTE
ANAKSTLVGHDTFTVKRPSAASIDAAKAGVNGNPDAVDD"
BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN
```

```
alignment_scores:
Quality: 846.00 Length: 164
Ratio: 5.190 Gaps: 0
Percent Similarity: 99.390 Percent Identity: 98.780
alignment_block:
US-09-589-870-2 x A93152
Align seg 1/1 to: A93152 from: 1 to: 1356
20 AlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAl 36
853 GCTGCCGACCTGCAGACCCGTCACAGACTCCAAAGCTCAGTTTCTGC 902
36 AAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGlySerThr 53
903 AGCCGAAGCTGTATCACTGCGACCTGTATTAACCACTGGGCTGCACT 952
53 heileValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGlySer 69
953 TCATTGTGACCGCTGCTGCGACGAGACTCTGACTGGCAGCTACGATCT 1002
70 AlaValGlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSe 86
1003 GCGGTTGTAAACGAGAAATCCCGCTACGTAAGTCTGAGCCCTTATGACTC 1052
86 rAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPrpValAlaT 103
1053 TGCACCTGCCACCGATGCTGCTGACCGCTGCGCTGCGACTGCTGCTT 1102
103 rPlyAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSerGlyGln 119
1103 GGAAGAAACAACTATCTATATCCGACAGCCGCACTACGTGCTGCCAA 1152
120 TyrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSe 136
1153 TACGTTGGCGGCTGCTGAGGCTCGTATCAACACTCAGTGGCTGTTAACTC 1202
136 rGlyThrThrGluAlaAsnAlaTrpPlySerThrLeuValGlyHisAspT 153
1203 CGGCACTACCGAAGCAATGCATGAAATCGACACTAGTAGTCAATGACA 1252
153 hrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLys 169
1253 CCTTACCAGAACTTAAGCCTTCTGCTGCTGATGATGCTGCCAAGAAA 1302
170 AlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
1303 GCAGGCGGTAAACAGCGTAACCTCTAGAGCTGTTCAGCAA 1344
seq_name: gb_pat1:AR082490
seq_documentation_block:
LOCUS AR082490 1356 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 4 from patent US 5973116.
ACCESSION AR082490
VERSION AR082490.1 GI:10009216
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos,A.Antoniou, Spooner,R,Anthony and Deonarain,M.
TITLE Compounds for targeting
JOURNAL Patent: US 5973116-A 4 26-OCT-1999;
FEATURES
source Location/Qualifiers
1..1356
/organism="unknown"
BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN
```

## alignment\_scores:

Quality: 846.00 Length: 164  
Ratio: 5.190 Gaps: 0  
Percent Similarity: 99.390 Percent Identity: 98.780

## alignment\_block:

US-09-589-870-2 x AR082490 ..

Align seg 1/1 to: AR082490 from: 1 to: 1356

```
20 Alaseralaseralasproserlyaspserysalaglnvalaser1 36
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
853 GCTGCCGACCTGCAGACCCGTCACAGAGCTCCAAGCTCAGGTTCTGC 902
36 aalaglualaglylthrlythrlythrlythrlythrlythrlythr 53
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
903 AGCGAAGCTGTATCATCTGCACCTGTATACCACTGGGCTGCACCT 952
53 heilevalthrlythrlythrlythrlythrlythrlythrlythr 69
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
953 TCATTGTACCGCTGCTGCGACGAGCTCTGACTGGCAGCTACGAATCT 1002
70 Alavalaglysalaglyserlythrlythrlythrlythrlythrly 86
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1003 GCGGTGTACGACAGATCCGCTAGCTACTGAGCGCTTATGACTC 1052
86 ralaprolathraspserlyserlythrlythrlythrlythrlythr 103
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1053 TGCACCTGCCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1102
103 rplyasnasnlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyr 119
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1103 GGAACAACTATCGTATATCGCACAGCCGCTACGTGCTGCGCAA 1152
120 TyrValaglylualaglylualaglylualaglylualaglylual 136
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1153 TACGTGGCGGTGCTGAGGCTCGTATCAACAGCTGAGCGCTTATCATC 1202
136 rlylthrlythrlythrlythrlythrlythrlythrlythrlythr 153
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1203 CGGACACACGAGAGCGATGATGAAATCGACACTAGTAGGTCTATGACA 1252
153 hrpethrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyr 169
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1253 CCTTACCAAGTTAAGCTTCTCTCTCTAGCATTTGCTGCCAGAAA 1302
170 Alaglyvalasnasnglyasnprouleuaspaalavalagln 183
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1303 GCAGGCGTAAACAGGTAACCTCTAGACGCTGTTCAGCAA 1344
```

seq\_name: gb\_pat1:A20698

seq\_documentation\_block:

LOCUS A20698 525 bp DNA PAT 22-AUG-1994  
DEFINITION fxa-strpa DNA sequence.  
ACCESSION A20698  
VERSION A20698.1 GI:583400  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct.  
artificial sequence.  
REFERENCE  
1 (bases 1 to 525)  
TITLE  
CARRIER-BOUND RECOMBINANT PROTEIN. PROCESS FOR PRODUCING IT AND ITS  
JOURNAL  
USE AS AN IMMUNOGEN AND VACCINE  
Patent: WO 91/3155-A 2 05-SEP-1991;  
FEATURES  
location/Qualifiers  
source  
1..525  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 128 a 140 c 138 g 119 t  
ORIGIN

## alignment\_scores:

Quality: 837.50 Length: 172  
Ratio: 5.045 Gaps: 1  
Percent Similarity: 96.512 Percent Identity: 95.349

## alignment\_block:

US-09-589-870-2 x A20698 ..

Align seg 1/1 to: A20698 from: 1 to: 525

```
12 Serleuthrthrvalserllethralseralasproserly 28
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1 TCTAGACTAGTGGATCCATC...GAGGTAAGTCTATGAGCCGCTCCAA 47
28 saspserysalaglnvalaserlaalaglylthrlythrlythrlythr 45
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
48 GCAGTCCAAAGCTCAGGTTTGTGACGCCGAGAGCTGTATCATCTGCACT 97
45 rplyasnglnleuglyserlythrlythrlythrlythrlythrlythr 61
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
98 GGTATTAACCAACTGGGCTGCACTTTCATTGTGACCGCTGGTGGCAGCA 147
62 Alaleuthrlythrlythrlythrlythrlythrlythrlythrlythr 78
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
148 GCTGTGACTGACCTTACGAAATCTGCGGTGTGTAACGACGAATCCGCTA 197
78 rlylthrlythrlythrlythrlythrlythrlythrlythrlythrly 95
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
198 CGTACTGACTGCGCGCTATGACTCTGCACCTGCCAGATGCGCTGTGTA 247
95 hrlylaleuglylthrlythrlythrlythrlythrlythrlythrly 111
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
248 CCGCTCTGGGCTGGAGCTGGCTTGAAGAAACAACTATCGTAATGCGCAC 297
112 Serlathrthrlythrlythrlythrlythrlythrlythrlythrly 128
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
298 AGCGCACCTACGTGCTGCTGCGCAATACGTTGGCGGTGAGGCTCTAT 347
128 easnlythrlythrlythrlythrlythrlythrlythrlythrlythr 145
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
348 CAACACTCAGTGGCTGTAAATTCGGGCACTACGAGACGAATGCCAAGA 397
145 ysserlythrlythrlythrlythrlythrlythrlythrlythrlythr 161
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
398 AATCGACACTAGTAGTCATGACACCTTTACCAAGTTAAGCTTCTGCT 447
162 Alaserlleaspaalalalyslysalaglyvalasnasnglyasnprou 178
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
448 GCTAGCATTTATGCTGCCAAGAAAGCAGCGGTAAACAGGTAACCTCT 497
178 uaspaalavalagln 183
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
498 AGACGCTGTTACAGCAA 513
```

seq\_name: gb\_pat1:I15647

seq\_documentation\_block:

LOCUS I15647 525 bp DNA PAT 02-APR-1996  
DEFINITION Sequence 2 from patent US 5470573.  
ACCESSION I15647  
VERSION I15647.1 GI:1250555  
KEYWORDS  
SOURCE  
ORGANISM  
Unknown.  
Unknown.  
REFERENCE  
1 (bases 1 to 525)  
TITLE  
Lubitz,W. and Szostak,M.P.  
Immunogens comprising the non-lytic membrane spanning domain of  
bacteriophages MS2 or PhiX174  
Patent: US 5470573-A 2 28-NOV-1995;  
FEATURES  
location/Qualifiers

source 1..525  
/organism="unknown"  
BASE COUNT 128 a 140 c 138 g 119 t  
ORIGIN

alignment\_scores:  
Quality: 837.50 Length: 172  
Ratio: 5.045 Gaps: 1  
Percent Similarity: 96.512 Percent Identity: 95.349

alignment\_block:  
US-09-589-870-2 x I15647 ..

Align seg 1/1 to: I15647 from: 1 to: 525

```
12 SerLeuThrThrValSerIleThrAlaSerAlaAspProSerLeu 28
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 TCTAGACTAGTGCATCCATC...GAGGGTAGGCTGATGACCGTCCAA 47
28 SAspSerIleValAlaGlnValSerAlaAlaGlnValIleThrGlyThr 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48 GGACTCCAAAGCTCAGGTTCTGACGCGAAGCTGTATCACTGCGACCT 97
45 rPtyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAspGly 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 GGTTATACCACTGGGGTGCAGCTTTCATTGTGACCGCTGTCGCGACGGA 147
62 AlaLeuThrGlyThrTyrGluSerAlaValAlaGlyAsnAlaGluSerArg 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 GCTCGACTGCGACCTACGATCTGCGGTGTGTAACGCAATCCCGCTA 197
78 rValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyT 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 CGTACTGCTGCGGCTTATGACTCTGACCTGCCACCGCATGCTCTGTGTA 247
95 hrAlaIleGlyTyrThrValAlaIleThrPlyAsnAsnTyrArgAsnAlaHis 111
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 CCGCTGCGGCTGCACTGCTGCGGCAATACCTGCGGTGCGATCGTAATGCGCAG 297
112 SerAlaThrThrTyrPheSerGlyGlnTyrValGlyAlaGlnAlaArgIle 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 AGCGCACTAGCTGCTGCGGCAATACCTGCGGTGCGATCGTAATGCGCAG 347
128 eAsnThrGlnIlePheLeuThrSerGlyThrThrGluAlaAsnAlaThrP 145
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
348 CAACACTCAGTGGCTGTATACATCCGCGCATACCGAAGCAATGATGATGGA 397
145 ySerThrLeuValGlyHisAspThrPheThrIleValIleProSerAla 161
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
398 AATGCACACTAGTAGTGCATGACCTTACCAAGTTAAGCTTCTGCT 447
162 AlaSerIleAspAlaAlaIleLysAlaGlyValAlaAsnAsnGlyAsnProLe 178
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
448 GCTAGCATTTGATGCTGCCAAGAAGACGACGCGTAAACAACGGAATACCTCT 497
178 uAspAlaValGlnGln 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
498 AGACCGCTGTCAGCA 513
```

seq\_name: gb\_pat1:109059

seq\_documentation\_block: 498 bp PAT 02-DEC-1994  
LOCUS 109059  
DEFINITION Sequence 38 from Patent WO 8809344.  
ACCESSION 109059  
VERSION 109059.1 GI:588242  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 498)  
AUTHORS Huston,J.S. and Oppermann,H.

TITLE TARGETED MULTIFUNCTIONAL PROTEINS  
JOURNAL Patent: WO 8809344-A 38 01-DEC-1988;  
FEATURES Location/Qualifiers  
source 1..498  
/organism="unknown"  
BASE COUNT 117 a 137 c 133 g 111 t  
ORIGIN

alignment\_scores:  
Quality: 835.00 Length: 162  
Ratio: 5.154 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.148

alignment\_block:  
US-09-589-870-2 x I09059 ..

Align seg 1/1 to: I09059 from: 1 to: 498

```
22 AlaSerAlaAspProSerLeuAspSerIleAlaGlnValSerAlaAlaGln 38
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 TCCGGTGGCGACCCGTCGCAAGACTCCAAAGCTCAGGTTCTGCTGCCGA 53
38 uAlaGlyIleThrGlyThrTyrPtyrAsnGlnLeuGlySerThrPheIleVal 55
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 AGCTGATATCACTGCGACCTGCTGATTAACCAACTGGGCTGCACTTTCATTG 103
55 aThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaVal 71
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 TGACCGCTGCTGGCGACGAGCACTGACTGCGACCTACGATCGAATCTCGGTT 153
72 GlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaPr 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 GGTACGCGAATCCCGCTACGACTGACTGCGCGGTATGACTCTGCACC 203
88 oAlaThrAspGlySerGlyThrAlaLeuGlyTyrThrValAlaThrPlyAs 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 TGCCACCGATGCTGCTGCTGACCGCTGCGGCTGCGACTGCTGGCTGGAAAA 253
105 snAsnTyrArgAsnAlaHisSerAlaThrThrTyrPheSerGlyGlnTyrVal 121
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 ACAACTATGCTAATGCCGACAGCCGCACTACGTGCTGCGCAATACGTT 303
122 GlyGlyAlaGlnAlaArgIleAsnThrGlnIlePheLeuThrSerGlyThr 138
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 GCGGCTGCTGAGGCTGCTGATCAACACTCAGTGGCTGTAAATCCGGGAC 353
138 rThrGluAlaAsnAlaThrPlySerThrLeuValGlyHisAspThrPheT 155
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 TACCGAAGCGAATGCTGGAATGACACTAGTAGTATGACACTTTTA 403
155 hrLysValIleProSerAlaAlaSerIleAspAlaAlaLysAlaGly 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 CCAAGTTAAGCTTCTGCTGCTAGCATTTGATGCTGCCAAGAAGACGAGC 453
172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454 GTAAACAACGGTTAACCTCTAGACGCTGTTCAGCA 489
```

seq\_name: gb\_pat1:A93838

seq\_documentation\_block: 498 bp DNA PAT 22-JAN-2000  
LOCUS A93838  
DEFINITION Sequence 7 from Patent WO9728263.  
ACCESSION A93838  
VERSION A93838.1 GI:6741969  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 498)  
AUTHORS Lubitz,W. and Sleytr,U.  
TITLE RECOMBINANT EXPRESSION OF S-LAYER PROTEINS

JOURNAL Patent: WO 9728263-A 07-AUG-1997;  
 JOURNAL LUBITZ WERNER (AT); SLEYER UWE (AT)  
 FEATURES Location/Qualifiers  
 source 1..498  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 120 a 137 c 131 g 110 t  
 ORIGIN

alignment\_scores:  
 Quality: 834.00 Length: 159  
 Ratio: 5.245 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-589-870-2 x A93838 ..

Align seg 1/1 to: A93838 from: 1 to: 498

```

25 Asproserlysaasprlysaalaglvalserialaagluaglytl 41
   |||||||
7  GACCCGTCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGAAAGCTGAT 56
   |||||||
41 eThrGlyThrTyrAsnGlnLeuGlySerThrPheIleValThrAlaG 58
   |||||||
57 CACTGGCACTGCTGATATACCAACTGGGGTGCAGCTTTCATTGTGACCGCTG 106
   |||||||
58 lYAlaAspGlyAlaLeuThrGlyThrGlySerAlaValAlaGlyAsnAla 74
   |||||||
107 GTGGGGAGGAGGCTCTGACTGCACTACGAATCTGCGGTTGGTAACCA 156
   |||||||
75 GluserArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
   |||||||
157 GAATCCCGCTACGACTGACTGCGCGCTTATGACTCTGCACCTGCCACCGCA 206
   |||||||
91 pGlySerGlyThrAlaLeuGlyTrrPrrValAlaAlaTrpLysAsnAsnTyrA 108
   |||||||
207 TGGCTGTGGTACCGCTCTGGGCTGAGCTGGCTTGGAAGAAACAATATC 256
   |||||||
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGlyAla 124
   |||||||
257 GTAATGGGCACAGCGCCACTGAGTCTGGCCAAATAGTGGGCGGTGCT 306
   |||||||
125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
   |||||||
307 GAGCGCTGTATCAACACTCAGTGGCTGTATACATCCGCACTACCGAAGC 356
   |||||||
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
   |||||||
357 GAATGCATGGAAATCGAACAAGTAGAGCATGACACCTTTACCAAGTTA 406
   |||||||
158 ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
   |||||||
407 AGCCTTCTGCTGCTAGCATGATGATCTCCAAAGAAACAGCGGTAACAC 456
   |||||||
175 GlyAsnProLeuAspAlaValGlnGln 183
   |||||||
457 GGTAACCTCTAGACGCTGTTCAGCAA 483
   |||||||
seq_name: gb_pat1:A00743

```

seq documentation block:  
 LOCUS A00743 507 bp DNA PAT 28-JAN-1993  
 DEFINITION S.avidinii synthetic gene for streptavidin.  
 ACCESSION A00743  
 VERSION A00743.1 GI:14605  
 KEYWORDS  
 SOURCE Streptomyces avidinii.  
 ORGANISM Streptomyces avidinii.  
 Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

AUTHORS Edwards, R.  
 TITLE SYNTHETIC GENE  
 JOURNAL Patent: WO 8903422-A 2 20-APR-1989;  
 JOURNAL British Bio-Technology Ltd  
 FEATURES Location/Qualifiers  
 source 1..507  
 /organism="Streptomyces avidinii"  
 /db\_xref="taxon:1895"  
 10..492  
 /note="synthetic gene"  
 /codon\_start=1  
 /transl\_table=11  
 /product="streptavidin"  
 /protein\_id="CAA0084.1"  
 /db\_xref="GI:14606"  
 /translation="MDPSKDSKAQVSAARAGITGTYNOIGSTFTYTAGACALIGTY  
 ESAVGNAESRYVLGRTDSAPATIDSGTALGWTVMKKNYRNAHSATTWGQYVSGAE  
 ARINTQMLTSGTTEANAMKSTLVGHDPTRKVPAAASIDAAKKAGVNNGNPLDAVOQ  
 "

BASE COUNT 125 a 137 c 130 g 115 t  
 ORIGIN

alignment\_scores:  
 Quality: 834.00 Length: 159  
 Ratio: 5.245 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-589-870-2 x A00743 ..

Align seg 1/1 to: A00743 from: 1 to: 507

```

25 Asproserlysaasprlysaalaglvalserialaagluaglytl 41
   |||||||
13  GACCCGTCAGAGACTCCAAAGCTCAGGTTCTGCAGCCGAAAGCTGAT 62
   |||||||
41 eThrGlyThrTyrAsnGlnLeuGlySerThrPheIleValThrAlaG 58
   |||||||
63 CACTGGCACTGCTGATATACCAACTGGGGTGCAGCTTTCATTGTGACCGCTG 112
   |||||||
58 lYAlaAspGlyAlaLeuThrGlyThrGlySerAlaValAlaGlyAsnAla 74
   |||||||
113 GTGGGAGCGAGGCTCTGACTGGCACTACGAATCTGGGTTGGTAACGA 162
   |||||||
75 GluserArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
   |||||||
163 GAATCCCGCTACGACTGACTGCGCGCTTATGACTCTGCACCTGCCACCGCA 212
   |||||||
91 pGlySerGlyThrAlaLeuGlyTrrPrrValAlaAlaTrpLysAsnAsnTyrA 108
   |||||||
213 TGGCTGTGGTACCGCTCTGGGCTGAGCTGGCTTGGAAGAAACAATATC 262
   |||||||
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGlyAla 124
   |||||||
263 GTAATGGGCACAGCGCCACTGAGTCTGGCCAAATAGTGGGCGGTGCT 312
   |||||||
125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
   |||||||
313 GAGCGCTGTATCAACACTCAGTGGCTGTAAACATCCGCACTACCGAAGC 362
   |||||||
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
   |||||||
363 GAATGCATGGAAATCGAACAAGTAGAGCATGACACCTTTACCAAGTTA 412
   |||||||
158 ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
   |||||||
413 AGCCTTCTGCTGCTAGCATGATGATCTCCAAAGAAACAGCGGTAACAC 462
   |||||||
175 GlyAsnProLeuAspAlaValGlnGln 183
   |||||||
463 GGTAACCTCTAGACGCTGTTCAGCAA 489
   |||||||

```



```

seq_name: gb_pat1:A00744
seq_documentation_block:
LOCUS      A00744             507 bp    DNA                PAT          28-JAN-1993
DEFINITION S.avidinii synthetic gene (reverse complement) for streptavidin.
ACCESSION   A00744
VERSION     A00744.1 GI:14607
KEYWORDS
ORGANISM    Streptomyces avidinii.
SOURCE      Streptomyces avidinii.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces
REFERENCE   1 (bases 1 to 507)
AUTHORS    Edwards, R.
TITLE       SYNTHETIC GENE
JOURNAL     Patent: WO 8903422-A 3 20-APR-1989;
            British Bio-Technology Ltd
COMMENT     #A00744 Is the reverse complement of #A00743.
FEATURES
            Location/Qualifiers
            source
BASE COUNT      115 a      130 c      137 g      125 t
ORIGIN
Alignment_scores:
Quality: 834.00           Length: 159
Ratio: 5.245              Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-589-870-2 x A00744/rev ..
Align seg 1/1 to reverse of: A00744 from: 1 to: 507
25 ASPPROSERLYASPSERLYSLALGIVALSERIALAAGTAACTGATC 41
495 GACCCTGCCTCAAGACTCCAAAGCTCAGATTCTTCACCCGAAGTGTTAT 446
41 ethrcglythrtprrpyraenglnleuglyserthrphelvalrphalag 58
|||||
445 CACTGTGGCACCTGTGTAATAACAAGTGGGGTGCACCTTATATGTGACCGCTG 396
58 IYLAASPGIYALALEUTHRCGLYTHRTPRGISERALAVALLGIYAASNALA 74
|||||
395 GTGGCGGACGAGAGCTCTGACTGGACACTACGATCTGGGGTTGGTAACGCCA 346
75 GLUSERARTGYRVALLLEUTHRGILYARGTYRASPSERIALAPROLATHRAS 91
345 GAATCCCGCTACGTACGTACGTGGCGCGTTATGACTCTGCACCTGGCACCGA 296
91 POLYSERCILYTHRALALEUGLYTRPTRLVVALATRPRLYASNASTRYTA 108
|||||
295 TGGCTCTGGTACCGCTCTGGGCTGGAGCTGGTGGTGA AAAACAATCATC 246
108 TGAASALAHISERIALATHRTHTRRPSESGLYGINTRYVALIGLYALAL 124
245 CTAAATGCCACACAGCCGCCACTACGTGCTGGGCCAATAACGTTGGCGGCT 196
125 GIUALAARGLILEASTHNGINTRPLEULEUTHRSERGILYTHRTRIUAL 141
|||||
195 GAGCGCTCTATCAACACATCACTGAGCTGTAAACATCCGCGACTACCGAAGC 146
141 AASNAIATHRRPLYSERTHLEUVALLGILYHNASPTHRPHETHLYSHAL 158
|||||
145 GAATGCATAGA AAATCGACACTAGTAGGCATAGACACCTTTAACCAAATTA 96
158 YSPROSERIALALASERILEASPMALALATLVSLSALAGIYVALASASN 174
95 ACCCTTTCGCTGCTACGATTCATCTGTGCAAGAAGAAGCGCGTAAACAC 46
175 GLYASNPROLEUASPALAVALLGNGLN 183

```

```

seq_name: gb_pat1:A39567

seq_documentation_block: 1296 bp DNA PAT 05-MAR-1997
LOCUS A39567 Sequence 6 from Patent WO9415644.
DEFINITION A39567
ACCESSION A39567
VERSION A39567.1 GI:2295849
KEYWORDS
SOURCE .
ORGANISM unidentified.
unidentified.
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 1296)
Epenetos/A.A./ Spooner,R.A. and Deonarin,M.
AUTHORS Epenetos/A.A./ Spooner,R.A. and Deonarin,M.
TITLE COMPOUNDS FOR TAPPING
JOURNAL Patent: WO 9415644-A 6 21-JUL-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication GB 2289679 951129.
FEATURES
source
1..1296
location/Qualifiers
1..1296
/organism="unidentified"
/db_xref="taxon:32644"
40..1287
/db_xref="taxon:32644"
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02450.1"
/db_xref="GI:2295850"
/translation="MKYLLEPTAAAGILLNAOPMAQVLOQBPAGELVYKPGASVKLSC
KASGITPTSTYMMHWKQRPGRGLEWIRIDPNSGGTIVYKSKATLLTYDKFSSTAY
MQLSLTSEDSAVYVCARVDYGSSTFDYWGCGTIVYSSGGSGGGGGGGGSGQAV
VTSEGLATTPGETVTLTKRSSTGAVTNTSYAMWVEKPDHLEPTGLIGTNNGAPGP
ARSGSLIDKKAALITGAQTEDEAIVFCALWYNNHWFGGTRKLVGLLEAPAAPA
DPSNDSKAOVSAAGAGTGTWYNOLGTFIVTGAQDALGTYESAIVGNRESRYLVLG
RYSAPAPTDSSGTRALGNTVWKNYNNRAHSATTWISGQYVGGALRINTQMLTSGTTE
ANMKSLVHDHFTKVPSSAAS"
BASE COUNT 312 a 342 c 347 g 295 t
ORIGIN
alignment_scores:
Quality: 745.00 Length: 144
Ratio: 5.210 Gaps: 0
Percent Similarity: 99.306 Percent Identity: 98.611
alignment_block:
US-09-589-870-2 x A39567 ..
Align seg 1/1 to: A39567 from: 1 to: 1296
20 AAlaserAlaSerAlaaspproSerLysaspsrLysaAlaGlnValSerAl 36
|||||
853 GCGTCCGCGACCTGCAGACCCGTCGCAAGGCTCAAGACCTCAGAGTTTCTGC 902
36 aAlaGluAlaGlyIleThrGlyThrTrpTyraGlnLeuGlySerThrp 53
|||||
903 ACCGGAAGCTGGATACACCTGGCACCTCGGATATACCAACATGGGGGTGCACTT 952
53 heIleValThrAlaGlyAlaaspglyAlaLeuThrGlyThrTrpGlnSer 69
|||||
953 TCATTCGTGACCGCTGCTGCGGACGGGCTCTACCTGCGCACCTACGCAATCT 1002
70 AAlaValGlyAsnAlaGlnSerArGTrpValIleuThrGlyArGTrpAspse 86
|||||
1003 GCGGTTGGTACGGCAGAAATCCGCGTACGACTGAGCCGTTATGACTC 1052
86 rAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThrValAlat 103
|||||
1053 TCGACCTGCGACCGATGGCTCGTACCGCTCTGGGCTGAGCTGAGCTT 1102
103 rPlyAsnAsnTrpArgAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
|||||

```

1103 GGAACCACTATCGTATCGGCACAGCCACTACGTGGTGGCCAA 1152  
120 TyrValGlyGlyAlaGluAlaArgIleasnThrGlnTrpLeuThrSe 136  
|||||  
1153 TACGTGGCGGCTGCTGAGGCTCGTATCAACAACAGTGGCTGTACATC 1202  
136 rGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspT 153  
|||||  
1203 CGGCACCTACCGACGAGCATGCAATCGAATCGACACTAGTACGTGATGACA 1252  
153 hrPheThrLysValLysProSerAlaAlaSer 163  
|||||  
1253 CCTTACCAAGTTAAGCCTTCTGCTGCTAGC 1284



```

1 ATGCGCAGATGCTGTGACGCCATCGCCGTTTCCTTCCCTGACCAAGCTCTC 50
PT r1leThrAlaSerAlaSerAlaAspProSerIysAspSerIysAlaGlny 34
XX |||||||
51 GATTACGGCCAGCCCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCCAAG 100
XX |||||||
34 a1SerAlaAlaGlnAlaGlnAlaGlnIleThrGlyThrTrpTyrAsnGlnIleuGly 50
XX |||||||
101 TCTCGCGCGCGCGAGCGCGCATCACCGGCACCTGGTACCAACACCTCGGC 150
XX |||||||
51 SerThrPheIleValThrAlaGlnAlaAspGlyAlaLeuThrGlyThrTy 67
XX |||||||
151 TCGACCTTCATCGTACCGCGCGCGCGCGCCCTCGACCGGACCTTA 200
XX |||||||
67 rglUserAlaValGlyAsnAlaGlnIleUserArgTyrValLeuThrGlyArgT 84
XX |||||||
201 CGAATCGGCGCTCGGCAACGCCGAGAGCGGCTACGCTCCGACCGGTCGT 250
XX |||||||
84 yraspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
XX |||||||
251 ACGACAGCGCGCGCGCCACCGAGCGAGCGCGACCGCCCTGGTGGAGC 300
XX |||||||
101 ValAlaTrpIysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
XX |||||||
301 GTGGCTGGAGGAATTAATCAACCGCACCCCTCCGCGACACGCTGGAG 350
XX |||||||
117 rglGlnTyrValGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 134
XX |||||||
351 CGGCACTACGTCGCGCGCGCGCGAGGAGATTAACACCCAGTGGCTGC 400
XX |||||||
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpIysSerThrLeuValGly 150
XX |||||||
401 TGACCTCGCGGACCAACCGAGGCCAAGCTGGAAGTCCACGCTGGTGGC 450
XX |||||||
151 HisAspThrPheThrIysValIysProSerAlaAlaSerIleAspAlaAl 167
XX |||||||
451 CACGACACCTTCACCAAGGTGAAGCCGTCGCGCCCTCCATCGACGCGGC 500
XX |||||||
167 alySLyAlaGlnAlaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
XX |||||||
501 GAACAAGCGCGCGCTCAACAAGCGCAACCCGCTGACGCGCTTCAAGCAG 549
XX |||||||

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:291073
seq_documentation_block:
ID 291073 standard; DNA; 563 BP.
XX
AC 291073;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomyces avidinii sps gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; sps gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomyces avidinii.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99MO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
WP1: 2000-195402/17.
XX
P-PSDB: Y80512.
DR

```

```

XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
XX
PS Examples; Page 84; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii sps gene as an example of the heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 563 BP; 107 A; 213 C; 169 G; 74 T; 0 other;

alignment_scores:
Quality: 936.00 Length: 183
Ratio: 5.115 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x 291073 ..
Align seg 1/1 to: 291073 from: 1 to: 563

1 MetArgIysIleValAlaAlaIleAlaValSerLeuThrValSe 17
|||
10 ATGGCGCAAGATCGCTGTGACGATCGCGCTTCCCTGACCAAGGCTCTC 59
|||
17 r1leThrAlaSerAlaSerAlaAspProSerIysAspSerIysAlaGlny 34
|||
60 GATTACGGCCAGCCCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCCAAG 109
|||
34 a1SerAlaAlaGlnAlaGlnAlaGlnIleThrGlyThrTrpTyrAsnGlnIleuGly 50
|||
110 TCTCGCGCGCGCGAGCGCGCATCACCGGCACCTGGTACCAACACCTCGGC 159
|||
51 SerThrPheIleValThrAlaGlnAlaAspGlyAlaLeuThrGlyThrTy 67
|||
160 TCGACCTTCATCGTACCGCGCGCGCGCGCCCTCGACCGGAACTTA 209
|||
67 rglUserAlaValGlyAsnAlaGlnIleUserArgTyrValLeuThrGlyArgT 84
|||
210 CGAATCGGCGCTCGGCAACGCCGAGAGCGGCTACGCTCCGACCGGTCGT 259
|||
84 yraspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
|||
260 ACGACAGCGCGCGCGCCACCGAGCGAGCGCGCACCGCCCTGGTGGAGC 309
|||
101 ValAlaTrpIysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
|||
310 GTGGCTGGAGGAATTAATCAACCGCACCCCTCCGCGACACGCTGGAG 359
|||
117 rglGlnTyrValGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 134
|||
360 CGGCACTACGTCGCGCGCGCGAGGAGATTAACACCCAGTGGCTGC 409
|||
134 eutHrSerGlyThrThrGlnAlaAsnAlaTrpIysSerThrLeuValGly 150
|||
410 TGACCTCGCGGACCAACCGAGGCCAAGCTGGAAGTCCACGCTGGTGGC 459
|||
151 HisAspThrPheThrIysValIysProSerAlaAlaSerIleAspAlaAl 167
|||
460 CACGACACCTTCACCAAGGTGAAGCGGCTCGCGCCCTCCATCGACGCGGC 509
|||

```

167 alyLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183  
|||||  
510 GAAGAGGCGCGCGTCACACACGCAACCGCCGTCGACCGCGTCGACGAG 558

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1997.DAT:T73193

seq\_documentation\_block:

ID T73193 standard; DNA; 638 BP.

XX T73193;

XX 27-APR-1998 (first entry)

XX wild-type streptavidin DNA.

XX Streptavidin: biotin; anti-interference reagent; detection: muteln;  
XX avidin: non-specific binding; ss.

XX Unidentified.

XX Key Location/Qualifiers  
FT CDS 50..601

FT sig\_peptide /tag- a  
FT /product- streptavidin

FT mat\_peptide /tag- b  
FT 122..598  
FT /tag- c  
FT /product- streptavidin

XX DE19637718-A1.

XX 02-OCT-1997.

XX 16-SEP-1996; 96DE-1037718.

XX 01-APR-1996; 96DE-1013053.

XX (BOE ) BOEHRINGER MANNHEIM GMBH.

XX Brandstetter H, Deger A, Engh R, Kopetzki E, Mueller R;  
XX Schmitt U;

XX WPI; 1997-482043/45.

XX P-PSDB: W29306.

XX Streptavidin and avidin muteins with reduced binding affinity for  
XX biotin - useful for reducing interference from nonspecific binding  
XX in assays

XX Disclosure; Page 16-17; 26pp; German.

XX This DNA sequence encodes a streptavidin which is used in a novel method  
XX of reducing interference from non-specific binding in assays. Muteins  
XX constructed from a core streptavidin or avidin sequence are selected that  
XX differ from the native polypeptide by at least one amino acid and have a  
XX binding affinity for biotin of less than 1010 l/mole. The biotin-bindable  
XX polypeptide may be present as a polymeric conjugate, e.g. with another  
XX polypeptide or protein, especially bovine serum albumin. These muteins  
XX are used as anti-interference reagents for reducing and/or avoiding  
XX nonspecific interactions in a process for detecting an analyte. In  
XX particular, they are used in assays where the streptavidin/avidin-biotin  
XX specific binding pair is involved for qualitative and/or quantitative  
XX determination of an analyte in a test sample, e.g. a heterogeneous  
XX immunoassay or a hybridisation assay. Despite having a lower binding  
XX affinity for biotin, the muteins have high immunological cross-reactivity  
XX with native streptavidin and avidin.

XX Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

alignment\_scores:

Quality: 936.00 Length: 183  
Ratio: 5.115 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

us-09-589-870-2 x T73193 ..

Align seg 1/1 to: T73193 from: 1 to: 638

1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSer 17  
|||||  
50 ATGCGCAAGATCGCTTGACACCATCCGCTTCCCTGACACCGCTCTC 99  
17 rleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34  
|||||  
100 GATTACGGCCAGCGCTTGCGACAGCCCTCCAGAGACTCGAAGGCGCCAGG 149  
34 aLseAlaAlaGlyAlaGlyIleThrGlyThrTPTyrAsnGlnLeuGly 50  
|||||  
150 TCTGCGCGCCGAGCGCGCATACCGCGCACCTGCTATACACAGCTCGGC 199  
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyTrrTy 67  
|||||  
200 TCGACCTTCATCTGACCGCGCGCGCGCGCGCTGACCGGAACTTA 249  
67 rGluSerAlaValGlyAsnAlaGlyIleSerArgTyrValLeuThrGlyArgT 84  
|||||  
250 CGAGTCGGCGCTCGGCAAGCGCGAGCGCGCTACCTCTGACCGGCTCTT 299  
84 yRAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrrPthr 100  
|||||  
300 ACGACAGCGCGCCGCGCGCACGCGCACGCGCGCTGCTGGTTGGACG 349  
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrrPse 117  
|||||  
350 GTGGCTGTGAAGAAATTACTACCGCAACGCCCATCTCGGACACAGTGGAG 399  
117 rGlyGlnTrpValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrrPleuL 134  
|||||  
400 CGGCGTGAAGTGTGGCGCGCGCGAGGATCAACACCGACGTCGTCG 449  
134 eutThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150  
|||||  
450 TGACCTCGGCGCACCGACGCGCAAGCGCTGGAAGTCCACGCTGGTCGC 499  
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167  
|||||  
500 CACGACACCTTCACCAAGGTGAAGCGCTCGCGCTCCATCGACGCGGC 549  
167 alyLysAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183  
|||||  
550 GAAGAGGCGCGCGTCACACACGCAACCGCCGTCGACCGCGTCGACGAG 598

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1998.DAT:V34714

seq\_documentation\_block:  
ID V34714 standard; DNA; 638 BP.  
XX V34714;  
XX 27-AUG-1998 (first entry)  
XX S. avidinii streptavidin cDNA.  
XX Streptavidin: ligand; binding affinity: mutant; isolation;  
XX purification; recover; immobilise; ss.  
XX Streptomycetes avidinii.  
XX OS  
XX FH Key Location/Qualifiers  
FT CDS 50..601  
FT /tag- a  
FT sig\_peptide 50..121

FT	mat_peptide	/*tag= b
FT		122.598
FT		/*tag= C
XX	/product= Streptavidin	
PN	EP835934-A2.	
PD	15-APR-1998.	
XX		
PE	09-OCT-1997;	97EP-0117504.
XX		
PR	10-OCT-1996;	96DE-1041876.
XX		
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX		
PI	Skeria A, Voss S;	
XX		
DR	WPI; 1998-218868/20.	
DR	P-PSDB; W59216.	
XX		
PT	Streptavidin mutants with higher binding affinity for peptide	
PT	ligands - have mutation in amino acid region 44-53, used to isolate,	
PT	purify or determine fusion proteins including these ligands	
XX		
PS	Disclosure; Page -: 21pp; German.	
XX		
CC	This sequence encodes a wild-type streptavidin protein isolated from	
CC	Streptomyces avidinii. This sequence is used to produce mutants which	
CC	are used in a method to assay the binding affinity of streptavidin	
CC	mutants. These mutants have a mutation within the amino acid (aa) region	
CC	44-53 of the wild-type protein show a higher binding affinity than the	
CC	wild-type for peptide ligands that include the sequence of formula	
CC	Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly,	
CC	or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can	
CC	be used to isolate, purify and determine proteins or to determine/recover	
CC	substances that contain streptavidin-binding groups. Such compounds may	
CC	also be used to immobilise fusions on microtitre plates, microbeads or	
CC	sensor chips.	
CC	NOTE: This sequence does not appear in the specification but is used to	
CC	make the mutant streptavidin sequence represented in V34715 and V34716.	
XX		
QQ	Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;	

```

alignment_scores:
    Quality: 936.00
    Ratio: 5.115
    Percent Similarity: 100.000
    Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x V34714
..

Align seg 1/1 to: V34714 from: 1 to: 638

1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrVal 17
|||||
50 ATGCCGAGATCGTCGTTGGACCATCGCCGTTCCCTGACCAAGGTC 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysASPSeLysAlaGln 34
|||||
100 GATTCAGCGCCAGCGGCTGGCGCAGACCCTCCAAAGACTCAAGGCCAG 149
|||||
34 aLSerAlaAlaGluAlaGlyrIleThrGlyThrTrpYrAsnGlnLeuGly 50
|||||
150 TCTGGCGCGCGAGGCGGCGCATCAACGGACACTGTGCAACCACTCGGC 199
|||||
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThr 67
|||||
200 TCGACCTTCATCGTAGACGCGGCGCGGCGAGCGGCCCTGACCGGAACCTA 249
|||||
67 rGluSerAlaValAlaLysAsnAlaGluSerArgTrpValLeuThrGlyArg 84
|||||
250 CGAGTCGGCCCTCGGCACAGCGCGAGACCGGTACGTCTCAACGCGTGT 299
|||||

```

```

      84 YRASPSSerAlaProAlaIthrAspGlySerGlyThrAlaIeuGlyTrpThr 100
      300 ACGACACGCGCCCGGCCACCGACGCGACGCGACCGCCTCGGTGGACG 349
      101 ValAlaTrpIlySasnaAsnTyrArgAsnAlaHisSerAlaThrThrTrpSe 117
      350 GTGGCCCTGGAGAAATTAATCTACCGCAACGCCCTCCGCGACCACTGGAG 399
      117 rGlyGlyTyrValGlyGlyAlaGlyAlaArgIleAsnThrGlyTrpLeu 134
      400 CGGCGACTACTCTCGCGCGCGCCGAGCGCAGATCAACACCCAGTGGCTGC 449
      134 eunThrSerGlyThrThrGlnAlaAsnAlaTrpIlySerThrIeuValGly 150
      450 TGACCTCTCGGACACACCGAGCGCAACGCTGGAACTCCACGCTGGCGAC 499
      151 HisAspThrPheThrIlyValIlyProSerAlaAlaIleAspAla 167
      500 CACGACACCTTCACCAAGGTGAAGCCGTCGCGCCTCATGTACGACGGCG 549
      167 aLySLysAlaGlyValAlaSnaGlyAsnProLeuAspAlaValGln 183
      550 GAAGAAAGCGCGGCTCAACAACGGCAACCCGCTCGACGCCGTTGACGACG 598

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT.z49867
seq_documentation_block:
ID      Z49867 standard; CDNA; 638 BP.
XX
AC      Z49867;
XX
DT      25-APR-2000 (first entry)

```

```

seq_documentation_block:
ID  Z49867 standard; cDNA; 638 BP.
XX
XX  Z49867;
XX
AC  Z49867;
XX
DT  25-APR-2000 (first entry)
XX
DE  Streptavidin gene for recombinant PART27 vector.
XX
KW  Potato, proteinase inhibitor-II; PI-II; streptavidin; worm;
KW  insect; plant-toxicous protein; pest resistance; moth; insect;
KW  grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;
KW  insecticidal; ss.
XX
XX  Synthetic.
XX
OS  Synthetic.
XX
FH  Key
FH  CDS
FT  50..601
FT  /*tag= a
FT  /product= "Streptavidin protein"
FT  sig_peptide
FT  50..121
FT  /*tag= b
FT  mat_peptide
FT  122..598
FT  /*tag= c
XX
XX  WO200004049-A1.
XX
XX  27-JAN-2000.
XX
XX  15-JUL-1999; 99WO-NZ00110.
XX
XX  15-JUL-1998; 98NZ-0331002.
XX
XX  (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
XX  Christeller JT, Sutherland PW, Murray C, Markwick NP, Philip BA;
XX  Malone LA, Burgess EPJ;
XX
XX  WPI: 2000-171244/15.
XX
XX  P-PSDB; Y44701.
XX
XX  New chimeric polypeptide and composition comprising the polypeptide
XX  useful for conferring pest resistance on plants
XX
XX  Disclosure; Fig 12; 11pp; English.
XX

```







seq\_name: /cgn2\_2/gcdata/geneseq/geneseqn/NA1993.DAT:Q53412

seq\_documentation\_block:

ID Q53412 standard; DNA; 552 BP.

AC Q53412;

DT 27-JUN-1994 (first entry)

DE Streptavidin gene.

KM Streptavidin; protein secretion; Bacillus subtilis; ss.

OS Streptomyces avidin11.

XX Key Location/Qualifiers

FT CDS 1..552

FT mat\_peptide /tag= a

FT sig\_peptide /tag= b

FT misc\_feature /tag= c

FT /tag= d

/note= "fused to npr signal peptide in pBE673"

PN W09324631-A.

XX 09-DEC-1993.

XX 27-MAY-1993; 93WO-US05240.

XX 29-MAY-1992; 92US-0891524.

XX (DUPO ) DU PONT DE NEMOURS & CO E. I.

XX Nagaraajan V;

XX WPI: 1993-405822/50.

XX P-PSDB; R44491.

XX Streptavidin prodn. from Bacillus subtilis - using signal protein  
XX from bacterial exo-protein and expression element from Gram  
XX positive bacterial protein.

XX Disclosure: Fig 1b; 54pp; English.

XX Tetrameric biologically active streptavidin is produced by secretion  
XX from Bacillus subtilis transformed with a plasmid encoding the  
XX sequence.

XX Sequence 552 BP; 105 A; 207 C; 167 G; 70 T; 3 other;

alignment\_scores:

Quality: 927.00

Ratio: 5.093

Percent Similarity: 99.454

Percent Identity: 99.454

alignment\_block:

US-09-589-870-2 x Q53412 ..

Align seg 1/1 to: Q53412 from: 1 to: 552

1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrValSe 17  
1 ATGGCGAAGATCGTCGTTCGACCATCGCGTTCCCTGACGAGGTC 50  
17 rIleThrAlaSerAlaAspProSerLysAspSerLysAlaGln 34  
51 GATTACGGCAGCGCTTCGCGAGAYMYCTCCAAAGACTCGAAGGCCAGG 100

seq\_name: /cgn2\_2/gcdata/geneseq/geneseqn/NA1998.DAT:V34715

seq\_documentation\_block:

ID V34715 standard; DNA; 638 BP.

AC V34715;

XX 27-AUG-1998 (first entry)

XX S. avidin11 streptavidin mutant #1 DNA.

XX Streptavidin; ligand; binding affinity; mutant; isolation;

XX purification; recover; immobilise; ss.

XX Streptomyces avidin11.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 50..601

FT sig\_peptide /tag= a

FT mat\_peptide /tag= b

FT /tag= c

FT /product= Streptavidin

FT mutation 251..262

FT /tag= d

FT /note= "Wild type GAGTCGGCCGCTC is replaced by

GTCAAGCGCGCT"

XX EP835934-A2.

XX 15-APR-1998.

XX 09-OCT-1997; 97EP-0117504.

XX 10-OCT-1996; 96DE-1041876.

XX

XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.  
 PA CGGCCAGTACCTCGCGCGCGAGCGAGATCAACACCCAGTGGCTGC 449  
 XX  
 PI Skerra A, Voss S;  
 XX  
 DR WPI: 1998-218868/20.  
 DR P-PSDB; W59217.  
 XX  
 PT Streptavidin mutants with higher binding affinity for peptide  
 PT ligands - have mutation in amino acid region 44-53, used to isolate,  
 PT purify or determine fusion proteins including these ligands  
 XX  
 PS Disclosure: Page -: 21pp; German.  
 XX  
 CC This sequence encodes a mutant streptavidin protein isolated from  
 CC Streptomyces avidinii where the residues ESAV at position 44-47 of  
 CC the mature wild type sequence are replaced by VVAR. This sequence is  
 CC used to produce mutants which are replaced in a method to assay the binding  
 CC affinity of streptavidin mutants. These mutants have a mutation within  
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher  
 CC binding affinity than the wild-type for peptide ligands that include  
 CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y  
 CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant  
 CC streptavidin mutants can be used to isolate, purify and determine  
 CC proteins or to determine/recover substances that contain  
 CC streptavidin-binding groups. Such compounds may also be used to  
 CC immobilise fusions on microtitre plates, microbeads or sensor chips.  
 CC NOTE: This sequence does not appear in the specification but has  
 CC been constructed from the wild-type streptavidin sequence represented  
 CC in V34714.  
 XX  
 SQ Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

alignment\_scores:  
 quality: 919.00 length: 183  
 ratio: 5.077 gaps: 0  
 Percent Similarity: 98.907 Percent Identity: 98.361

alignment\_block:

US-09-589-870-2 x V34715 ..

Align seg 1/1 to: V34715 from: 1 to: 638

1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSe 17  
 |||||  
 50 ATGCGCAGAGATCGTGTGACGACATCGCTTCCTGACACGAGTCTC 99  
 17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnv 34  
 |||||  
 100 GATTACGGCCAGCGCTTCGGCAGACCCCTCCAAAGACTCCGAAGCCACAG 149  
 34 aISerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGly 50  
 |||||  
 150 TCTCGCGCGCGGAGCGGCATCACCGGCACTGGTACAACACGACTCGGC 199  
 51 SerThrPheIleValThrAlaGlyAlaAspGlyValAlaLeuThrGlyTrp 67  
 |||||  
 200 TCGACCTTCATCTGACCGCGGCGGCGGCGGCGCTTCGACCGGAACCTTA 249  
 67 rGluSerAlaValGlyValAsnAlaGluSerArgTyrValIleuThrGlyArgT 84  
 |||||  
 250 CGTACCGCGCGCGTGGCAACGCCGAGACCGCTACGTCTACCGGTCGT 299  
 84 TyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrPThr 100  
 |||||  
 300 ACACACACCGCGCGCGCACCGACGCGACCGCGCCCTCGGTTGGACG 349  
 101 ValAlaIleThrLysAsnTrpTyrArgAsnAlaHisSerAlaThrThrTrpSe 117  
 |||||  
 350 GTGGCCCTGGAGAAATTAATACCGCAAGCCCACTCCGCGACCAACGTTGAG 399  
 117 rGlyGlnTyrValGlyValAlaGluAlaArgIleAsnThrGlnTrpLeuL 134

|||||  
 400 CGGCCAGTACCTCGCGCGCGAGCGAGATCAACACCCAGTGGCTGC 449  
 134 eunrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150  
 |||||  
 450 TGACCTCGGACACCGAGCGCAACGCGCTGAGTCCAGCTGTGGCG 499  
 151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167  
 |||||  
 500 CACGACACCTTCACCAAGGTGAAAGCCGTCGCCCTCCATCGACGCGGC 549  
 167 aLysLysAlaGlyValAlaAsnGlnLysAsnProLeuAspAlaValGlnGln 183  
 |||||  
 550 GAACAAGCGCGCGCTTCACAACAGCGCAACCCCTCGACGCGCTTACGACG 598  
 seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1998.DAT:V34716  
 seq\_documentation\_block:  
 ID V34716 standard; DNA: 638 BP.  
 XX  
 AC V34716;  
 XX  
 DT 27-AUG-1998 (first entry)  
 XX  
 DE S. avidinii streptavidin mutant #2 DNA.  
 XX  
 KW Streptavidin; ligand; binding affinity; mutant; isolation;  
 KW purification; recover; immobilise; ss.  
 XX  
 OS Streptomyces avidinii.  
 OS Synthetic.  
 XX  
 FH Key location/Qualifiers  
 FT CDS 50..601  
 FT /\*tag= a  
 FT sig\_peptide 50..121  
 FT /\*tag= b  
 FT mat\_peptide 122..598  
 FT /\*tag= c  
 FT /\*product= Streptavidin  
 FT mutation 251..262  
 FT /\*tag= d  
 FT /\*note= "Wild type GAGTCGCGCTC is replaced by  
 ATCGTGCAGAG"  
 FT  
 XX  
 PN EP835934-A2.  
 XX  
 PD 15-APR-1998.  
 XX  
 PF 09-OCT-1997; 97EP-0117504.  
 XX  
 PR 10-OCT-1996; 96DE-1041876.  
 XX  
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.  
 XX  
 PI Skerra A, Voss S;  
 XX  
 DR WPI: 1998-218868/20.  
 DR P-PSDB; W59218.  
 XX  
 PT Streptavidin mutants with higher binding affinity for peptide  
 PT ligands - have mutation in amino acid region 44-53, used to isolate,  
 PT purify or determine fusion proteins including these ligands  
 XX  
 PS Disclosure: Page -: 21pp; German.  
 XX  
 CC This sequence encodes a mutant streptavidin protein isolated from  
 CC Streptomyces avidinii where the residues ESAV at position 44-47 of  
 CC the mature wild type sequence are replaced by IGAH. This sequence is  
 CC used to produce mutants which are replaced in a method to assay the binding  
 CC affinity of streptavidin mutants. These mutants have a mutation within  
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher  
 CC binding affinity than the wild-type for peptide ligands that include

CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y  
 CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant  
 CC streptavidin mutants can be used to isolate, purify and determine  
 CC proteins or to determine/recover substances that contain  
 CC streptavidin-binding groups. Such compounds may also be used to  
 CC immobilize fusions on microtitre plates, microbeads or sensor chips.  
 CC NOTE: This sequence does not appear in the specification but has  
 CC been constructed from the wild-type streptavidin sequence represented  
 CC in V34714.

XX  
 XX Sequence 638 BP; 116 A; 242 C; 194 G; 86 T; 0 other;

# alignment\_scores:

Quality: 917.00 Length: 183  
 Ratio: 5.066 Gaps: 0  
 Percent Similarity: 98.907 Percent Identity: 98.361

# alignment\_block:

US-09-589-870-2 x V34716 ..

Align seg 1/1 to: V34716 from: 1 to: 638

```

1 MeArGtYsIleValAlaAlaIleAlaValSerLeuThrValSe 17
|||||
50 ATGGCGAAGATCGTGTGACGACCGCTTCCCTGACGACGCTCTC 99
|||||
17 rIleThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnV 34
|||||
100 GATTACCGCGCAGCTTGGCAGACCCCTCCAGGACTCGAAGGCCAGG 149
|||||
34 AlSerAlaAlaGlnAlaGlyIleThrGlyThrPTyrAsnGlnLeuGly 50
|||||
150 TCTCGGCGCGGAGCGCGCATCACCGGACCTGTACTCAACACCTCGGC 199
|||||
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTy 67
|||||
200 TCGACCTTCATCGTACCGCGCGCGCGCGCGCGCTGACCGGACCTTA 249
|||||
67 rGluSerAlaValAlaGlyAlaGluSerArgTyrValLeuThrGlyArgT 84
|||||
250 CATGCGTGGAGGAGCAACCGCGAGACCGCTGACCTGCTGACCGTCTT 299
|||||
84 yRapASerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
|||||
300 ACGACACGCGCGCGCGCACCGACGCGCGCGCGCGCTGGTGGACG 349
|||||
101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrpPse 117
|||||
350 GTGGCTCGAAGAAATTAACCGCAAGCCCACTCCCGACCACTGGAG 399
|||||
117 rGlyGlnTyrValGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuL 134
|||||
400 CGGCGAGTACGTCGCGCGCGCGAGGAGATCAACACCCAGTGGCTGC 449
|||||
134 eutThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGly 150
|||||
450 TGACCTCCGGCACCAACGAGCCAGCTGGAGATCCACGCTGCTGCGC 499
|||||
151 HisAspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAl 167
|||||
500 CACGACACCTTCACCAAGGTGAAGCCGCTCCGCGCTTCATCGACGCGGC 549
|||||
167 AlYsLysAlaAlaGlyValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
550 GAAGAAGGCGCGGCTCAACAACGCAACCGCGCTGACGCGCTTGACGAG 598
|||||

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1994.DAT:Q70659

seq\_documentation\_block:

ID 070659 standard; DNA; 1356 BP.

XX  
 AC 070659;

XX  
 DT 26-MAR-1995 (first entry)

XX  
 DE ScFv PRAS108 and PRAS112 DNA.

XX  
 KW Amplification; single chain variable region fusion protein; PCR; ss.

XX  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers

FT CDS 40..1348

FT FT /\*tag= a

XX  
 PN M09415644-A.

XX  
 PD 21-JUL-1994.

XX  
 PF 17-JAN-1994; 94MO-GB00087.

XX  
 PR 15-JAN-1993; 93GB-0000686.

XX  
 PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.

XX  
 PI Deonaraia M, Epenetos AA, Spooner RA;

XX  
 DR WPI: 1994-248907/30.

XX  
 DR P-PSDB; R56483.

XX  
 PS New cpds, comprising a targeting portion and a cytotoxic portion

XX  
 CC - used esp. for treating mammals for destroying target cells,

XX  
 CC partic. tumour cells

XX  
 CC Claim 36; Fig 9; 114pp; English.

XX  
 CC The sequence is that of the scFv PRAS108 and PRAS112 between HindIII and

XX  
 CC EcoRI site obtd. by PCR.

XX  
 CC See also Q70652-69.

XX  
 SQ Sequence 1356 BP; 332 A; 359 C; 359 G; 306 T; 0 other;

alignment\_scores:  
 Quality: 846.00 Length: 164  
 Ratio: 5.190 Gaps: 0  
 Percent Similarity: 99.390 Percent Identity: 98.780

# alignment\_block:

US-09-589-870-2 x Q70659 ..

Align seg 1/1 to: Q70659 from: 1 to: 1356

```

20 AlASerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAl 36
|||||
853 GCTGCGGACCTGCGAGACCGCTCCAGAGACTCCAAAGCTCAGTTTCTGC 902
|||||
36 AlAlaGlnAlaGlyIleThrGlyThrPTyrAsnGlnLeuGlySerThrP 53
|||||
903 AGCGAAGCTGTATCACTGACCTGTATACCACTGGGGTGCACCTT 952
|||||
53 heIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGluSer 69
|||||
1053 TCATTGTGACCGCTGTGTGGAGCGAGCTGTGACTGGCGACCTACGAACT 1002
|||||
70 AlAlaValGlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSe 86
|||||
1003 GCGGTTGTTAAGCGAGATCCCGCTAGCTGACTGCGCGCTTATGACTC 1052
|||||
86 rAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThrValAlaTr 103
|||||
1053 TGCACCTGCGCACCGATGCTCTGTGATACCGCTGTGGGCTGAGCTGTGCTT 1102
|||||
103 rPLysAsnAsnTyrArgAsnAlaHisSerAlaThrTrpSerGlyGln 119
|||||

```

```

1103 GGAACAACTATGCTATGCGCACAGCCGCACTACGTGCTGGCCAA 1152
120 TyrValGlyGlyAlaGluAlaArgIleasnThrGlnTrpLeuThrSe 136
1153 TACGTTGGCGGCGCTGAGGCTGCTATCAACACTGAGGCTGTAAATC 1202
136 rGlyThrThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspT 153
1203 CGGCACACTACGGAAGCATGATGGAATCGACACTAGTAGTCAATGACA 1252
153 hrpPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLys 169
1253 CTTTACCAAGATTAAAGCTTCTGCTGCTACGATTGATGCTGCCAAGAA 1302
170 AlaGlyValaAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
1303 GCAGGCGTAAACAAAGCTAACCTTATACACGCTGTTCAGCAA 1344

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA2000.DAT:291076
seq_documentation_block:
ID 291076 standard; DNA; 495 BP.
XX
AC 291076;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomyces avidinii mprot gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KM viability; mprot gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomyces avidinii.
XX
PN MO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
DR WPI: 2000-195402/17.
XX
DR P-PSDB; Y80515.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 86-87; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii mprot gene as an example of a heterologous gene introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 495 BP; 98 A; 188 C; 151 G; 58 T; 0 other;

```

```

alignment_scores:
Quality: 838.00 Length: 160
Ratio: 5.237 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-589-870-2 x 291076 ..

Align seg 1/1 to: 291076 from: 1 to: 495

24 AlaaspProSerLysaspSerLysAlaGlnValSerAlaAlaGluAlaGly 40
11 GCTTACCCCTCCAAAGACTCGAAGGCCAGAGTCTGCGCGCCGAGGCGCG 60
40 YIleThrClYThrTrpPyrAsnGlnLeuGlySerThrPheIleValIhrra 57
61 CATCACGCGCACGCTGTAACAACACAGCTCGGCTCACTTCATCGTACCG 110
57 laGlyAlaAspGlyAlaLeuThrGlyThrTrpGlyGlnSerAlaValGlyAsn 73
111 CGGCGCGCGAGCGCGCCCTGACCGGAACCTACGAGTGGCCGCTGGCAAC 160
74 AlaGlnSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaTh 90
161 GCCGAGAGCCGCTACGTCCTGACCGGTCGTTACGACAGCGCGCCGCGCAC 210
90 rAspGlySerGlyThrAlaLeuGlyTTPThrValAlaTrpLysAsnAsnT 107
211 CGAGCGAGCGCGCACGCGCTCGGTGACGCGTGGCAAGAAATTAAT 260
107 YrArgAsnAlaHisSerAlaThrTrpSerGlyGlnTyrValGlyGly 123
261 ACCGCAAGCCCACTCCGCGACACAGCTGAGCGCCGACGTCGCGCGCC 310
124 AlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrG 140
311 GCCGAGCGAGGATCAACACCGATGGCTGCTGACCTCGCGCACACCGA 360
140 uAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLys 157
361 GGCACACGCTCGGAAGTCAACGCTGGGCGCACGACACCTTCACCAAGG 410
157 aLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValaAsn 173
411 TGAAGCGCTCGCGCCCTCATCGACGCGCGGAAGAGCGCGCGCTCAAC 460
174 AsnGlyAsnProLeuAspAlaValGlnGln 183
461 AACGCAACCCGCTCGACGCGCTTCAGACAG 490

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA2000.DAT:291075
seq_documentation_block:
ID 291075 standard; DNA; 492 BP.
XX
AC 291075;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomyces avidinii mprot gene.
XX
KW Plant somatic tissue degeneration; plant essential factor; depletion;
KM viability; mprot gene; plant development; plant morphology; flower;
KW fruit plant; ds.
XX
OS Streptomyces avidinii.
XX
PN MO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX

```

```
PR 03-AUG-1998; 98IL-0125632.
XX
XX (AGRI-) AGRIC RES ORG.
XX
XX Kapulnik Y, Ginzberg I;
XX
XX WPI: 2000-195402/17.
XX
XX P-PSDB; Y80514.
XX
XX Degeneration of somatic plant tissue by expression of a heterologous
XX protein, useful for controlling plant development and morphology, such
XX as decreasing the number of flowers present to increase the number of
XX fruit -
XX
XX Examples; Page 86; 91pp; English.
XX
XX The invention relates to a method of effecting degeneration of a somatic
XX plant tissue by expressing a heterologous protein capable of binding a
XX plant essential factor (PEF), in somatic plant tissue cells, where
XX heterologous protein expression causes depletion of the PEF so the plant
XX viability is maintained, while simultaneous degeneration of the somatic
XX plant tissue is effected. This sequence represents the Streptomyces
XX avidin1 prot gene as an example of a heterologous gene introduced into
XX the plants. The methods can provide for the selective and optionally
XX reversible cell degeneration in somatic plant tissue. They can be used
XX for artificially controlling plant development and morphology. They can
XX be used e.g. to decrease the number of flowers in fruit producing plants
XX so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 492 BP; 95 A; 188 C; 151 G; 58 T; 0 other;

alignment_scores:
Quality: 835.00 Length: 161
Ratio: 5.186 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.758

alignment_block:
US-09-589-870-2 x Z91075 ..

Align seg 1/1 to: Z91075 from: 1 to: 492

23 SerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnVal 39
:.....:
5 GCAGTTGACCCCTCCAGAGACTCGAAGGCCCAAGGCTCGGCGCCGCGAGGC 54

39 agLyleThrGlyThrTrpTyraSngInLeuGlySerThrPheIleValT 56
|||||
55 CGGCATACGGGCACCTGTAACAACGACGCTGGCTGACCTTCAATCGTGA 104

56 hralaGlyAlaAspGlyAlaLeuThrGlyThrTyrgLuserAlaValGly 72
|||||
105 CCGCGGGCGCGGACGCGCCCTGACCGGAACCTACGAGTGGCGCTGGCG 154

73 AsnAlaGlnSerArgTyraIleuThrGlyArgTyraSspSerAlaProAl 89
|||||
155 AAGCGCGAGCGCGCTACGTCGTACCGGTGTTACACAGCGCCCGCGG 204

89 aThrAspGlySerGlyThraIleuGlyTrpThraIaIaTrpLysAsna 106
|||||
205 CACGACGGGAGCGGCGCCCTCGGTGGACGCTGGCTCGAAGATA 254

106 snTyraIrgaAsnaIaHisSerAlaThrThrTrpSerGlyInTyraValGly 122
|||||
255 ACTAACCCCAAGCCCACTCCGCGACACGCTGAGCGCGCGACGTAAGTGGC 304

123 GlYAlaGlnAlaArgIleAsnThrGlnTrpLeuIleuThrSerGlyThrTh 139
|||||
305 GCGCGCCGAGGAGATCAACCCCAAGTGGCTGTGTACCTCGGCGACACAC 354

139 rGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrL 156
|||||
355 CGAGGGCCAAAGCCGTGGAAGTCCACGCTGGTGGCGCACACCTTCACCA 404
```

```
156 ysValLySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyVal 172
|||||
405 AGTGAAGCCGTCCGCGCCCTCCATCGACGCGCGGCAAGAGCGCGGCTC 454

173 AsnAsnGlyAsnProLeuAspAlaValGlnGln 183
|||||
455 AACACGGCAACCCGCTCGACGCGCTTCACAGCAG 487

seq_name: /cgn2_2/gcdata/geneseq/geneseqn/NA1988.DAT:N80186
seq_documentation_block:
ID N80186 standard; DNA; 498 BP.
XX
XX N80186;
XX
XX 13-OCT-1990 (first entry)
XX
XX DE Encodes biosynthetic protein with streptavidin trailer.
XX
XX KW Biosynthetic Antibody Binding site (BABs); variable domain;
XX anti-CEA monoclonal antibody; streptavidin; ss.
XX
XX OS Synthetic.
XX
XX PN WO8809344-A.
XX
XX PD 01-DEC-1988.
XX
XX PF 19-MAY-1988; 88WO-US01737.
XX
XX PR 21-MAY-1987; 87US-0052800.
XX
XX PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
XX PI Huston JS, Oppermann H;
XX
XX DR WPI: 1988-353928/49.
XX
XX DR P-PSDB; P80160.
XX
XX PT Recombinant multifunctional protein -
XX having an antibody binding site and a sequence for biological activity,
XX for ion sequestering or binding to a solid support.
XX
XX PS Disclosure; ; 115pp; English.
XX
XX CC Encodes multi-functional biosynthetic protein comprising single
XX chain BABs and streptavidin protein trailer linked via a spacer
XX sequence.
XX See also N80171-N80192.
XX
SQ Sequence 498 BP; 117 A; 137 C; 133 G; 111 T; 0 other;

alignment_scores:
Quality: 835.00 Length: 162
Ratio: 5.154 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.148

alignment_block:
US-09-589-870-2 x N80186 ..

Align seg 1/1 to: N80186 from: 1 to: 498

22 AlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGln 38
:.....:
4 TCCGCTGGCAGACCCGCTCCAGAGACTCCAAAGCTTCTGCTGCCGA 53

38 uAlaGlyIleThrGlyThrTrpTyraSngInLeuGlySerThrPheIleV 55
|||||
54 AGCTGTATGACTGGCACCTGTAACCAACTGGGCTGCACCTTTCATTG 103

55 alThraIaGlyAlaAspGlyAlaLeuThrGlyThrTyrgLuserAlaVal 71
```

```

|||||
104 TGACCGGTGGTGGACGAGCTGACTGCGACCTAGCAATCTGGCGTT 153
72 GLYsnaIaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaPr 88
154 GGTACCGAGATCCCGTACGTACTAGCTGGCGGTATGACTCTGAC 203
88 AlaThrAspGlySerGlyThrAlaLeuGlyTyrThrValAlaTrpLysA 105
204 TGGCACCAGTGGCTGCTGATCCGCTGCGGCTGAGACTGTGCTGGAAAA 253
105 sNaSnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrVal 121
254 ACAACTATCGTAATGCCACAGCCGCACTACGTGCTGGCCAAATACGTT 303
122 GlyGlyAlaGluAlaArgTyrLeaSnThrGlnTrpLeuLeuThrSerGlyTh 138
304 GCGGCTGCTGAGCTGCTATCAACACACAGTGGCTGTTAACATCGGCGAC 353
138 rThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheT 155
354 TACGGAAGCGATGATGGAATGACACACTAGTGTGATGACACCTTTA 403
155 hrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGly 171
404 CCAAGGTTAAGCCTTCTGCTGCTAGCATGTGATGCTGCCAAGAAAGCAGC 453
172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
454 GTAAACAACGTAACCTCTAGACGCTGTTCAGCAA 489

seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT:T75491
seq_documentation_block:
ID T75491 standard; DNA; 498 BP.
XX
XX T75491;
XX
XX 27-FEB-1998 (first entry)
XX
XX DNA for streptavidin.
XX
XX S-layer; sbs-A; vaccine; adjuvant; carrier; hybridisation assay;
XX molecular spinning nozzle; molecular laser; streptavidin; ds.
XX
XX Streptomyces avidinii.
XX
XX DE19603649-A1.
XX
XX 07-AUG-1997.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX 01-FEB-1996; 96DE-1003649.
XX
XX (LUBI/) LUBITZ W.
XX (SLEY/) SLEYTR U.
XX
XX Kuen B, Lubitz W, Sleytr U;
XX
XX WPI; 1997-394558/37.
XX
XX Preparation of S-layer proteins by expressing sbs-A gene in Gram
XX negative bacterium - or new sbs-B gene in any host, also new
XX recombinant proteins containing heterologous inserts, e.g.
XX epitope(s), useful as vaccines and adjuvants
XX
XX Example 8; Page 26; 31pp; German.
XX
XX S-layer structures comprising the novel Bacillus stearothermophilus
XX PV72 S-layer protein, sbs-A, can be used as vaccines or adjuvants,
XX particularly when they include a bacterial ghost that may contain
XX additional epitopes in its membrane. Other uses of recombinant

```

```

CC sbs-A, depending on the nature of the inserted peptide, are as an
CC universal carrier for biotinylated reactants for use in
CC immunological or hybridisation assays (the insert is streptavidin,
CC preferably encoded by the present sequence), to induce immune
CC responses (epitopes), as a reagent for removing cytokine or toxin
CC from serum (antigenic epitopes), as a molecular spinning nozzle
CC (polyhydroxybutyrate synthase) and as a molecular laser
CC (luciferase).
XX
XX Sequence 498 BP; 120 A; 137 C; 131 G; 110 T; 0 other;

alignment_scores:
    Quality: 834.00      Length: 159
    Ratio: 5.245        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-589-870-2 x T75491 ..

Align seg 1/1 to: T75491 from: 1 to: 498

25 AspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGlyI 41
7 GACCCGTCGAAGAGCTCCAAAGCTCAGTTCTGCGACCGAAGTGCTAT 56
41 eThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValIThrAla 58
57 CACTGGCACCCTGGTATTAACCAACCTGGGGTGCACCTTCATTGTAACCCCTG 106
58 lYAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaValGlyAsnAla 74
107 GTGCGGACGAGAGCTCTAGCTGCGACCTACGAATCTGGGTGGTAACGA 156
75 GluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAs 91
157 GAATCCCGCTAGCTAGTACTGGCGGCTTATGACTCTGCACCTGCCACCGA 206
91 pGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyrA 108
207 TGGCTCTGTACCGCTCTGGGCTGGAGCTGTGGCTGGAAAAACAACATATC 256
108 rGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyrValGlyGlyAla 124
257 GTATGCGCACAGCGCCACTACGTGCTGCTGCCAATACGTTGGCGGTGCT 306
125 GluAlaArgTyrLeaSnThrGlnTrpLeuLeuThrSerGlyThrThrGluAl 141
307 GAGGCTCGTATCAACACTCAGTGGCTGTTAAACATCCGGCACCTCCGAGC 356
141 aAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValL 158
357 GAATGATGGAAATCGACACTAGTGTGATGATGACACCTTTCACCAAGTTA 406
158 ySProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsn 174
407 AGCCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 456
175 GlyAsnProLeuAspAlaValGlnGln 183
457 GTTACCCCTCTAGACGCTGTTCAGCAA 483

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 5, 2001, 21:15:35 ; Search time 62.3 Seconds  
(without alignments)  
199.451 Million cell updates/sec

Title: US-09-589-870-2  
Perfect score: 936  
Sequence: 1 MKRIVVAIAIVSLTIVSITA.....IDAKKAGVNNGNPLDAVQ 183

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 segs, 67900655 residues  
Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	936	100.0	183	2	A23513	streptavidin precu
2	932	99.6	183	2	S57284	streptavidin v1 pr
3	897	95.8	183	2	S57285	streptavidin v2 pr
4	146.5	15.7	150	2	S42204	avidin-related pro
5	143.5	15.3	152	1	VICH	avidin precursor [
6	121.5	13.0	150	2	S42201	avidin-related pro
7	120.5	12.9	150	2	S42203	avidin-related pro
8	120	12.8	1064	2	A40136	fibropellin Ia - s
9	116.5	12.4	150	2	S42202	avidin-related pro
10	112	12.0	595	2	A48658	flagellin - Escher
11	101	10.8	595	2	B48658	flagellin - Escher
12	100.5	10.7	1026	2	A48895	paracrystalline su
13	99.5	10.6	1477	2	B43855	high-molecular-wel
14	99.5	10.6	2468	2	AB3412	hypothetical prote
15	96.5	10.3	741	2	F70722	probable g1cb prot
16	94.5	10.1	507	2	T35677	probable hydrolase
17	94	10.0	570	2	S56132	cellulase (EC 3.2
18	94	10.0	967	2	S66852	hypothetical prote
19	93	9.9	1225	1	B64234	hypothetical prote
20	92.5	9.9	1018	2	H83135	probable adhesin p
21	92	9.8	1217	2	S52714	sericin1B - silkw
22	91.5	9.8	1122	2	G64887	probable tail fibre
23	91.5	9.8	2249	2	A41477	190k surface anti
24	91.5	9.7	867	2	T45463	membrane glycoprot
25	90.5	9.7	2761	2	T21064	hypothetical prote
26	90	9.6	13288	2	T03099	mucin, submaxillar
27	89	9.5	255	1	CPSMUU	muramoylpentapepti
28	89	9.5	488	2	A32235	poly(1,3-hydroxybuty
29	89	9.5	866	2	T45462	membrane glycoprot

30	88.5	9.5	465	2	A47023	S-layer protein -
31	87.5	9.3	457	2	A53669	streptomycies grise
32	87	9.3	666	2	A42296	lysozyme 2 (EC 3.2
33	86	9.2	598	2	T38403	probable nucleopor
34	86	9.2	605	2	T33913	hypothetical prote
35	85.5	9.1	602	2	T45278	oligopeptide ABC t
36	85	9.1	303	2	T35616	probable membrane
37	85	9.1	528	2	S69589	hypothetical prote
38	85	9.1	757	1	VGBEX1	glycoprotein X pre
39	85	9.1	872	2	S49541	cellulase - Cellul
40	85	9.1	1335	2	T30211	autolysin E - Stap
41	85	9.1	1536	2	A43855	high-molecular-wel
42	84.5	9.0	962	2	S03818	carboxymethylcellu
43	83.5	8.9	414	2	A29835	SalI protein - Esc
44	83.5	8.9	490	2	A24262	phase-1 I flagellin
45	83.5	8.9	507	2	T50398	hypothetical serin

## ALIGNMENTS

RESULT 1  
A23513  
streptavidin precursor - Streptomyces avidin1  
C:Species: Streptomyces avidin1  
C>Date: 03-Nov-1987 #sequence\_rev: 03-Nov-1987 #text\_change 10-Dec-1999  
C:Accession: A23513; S11540  
R:Argarana, C.E.; Kuntz, I.D.; Birken, S.; Axel, R.; Cantor, C.R.  
Nucleic Acids Res. 14, 1871-1882, 1986  
A:Title: Molecular cloning and nucleotide sequence of the streptavidin gene.  
A:Reference number: A23513; MID:86148514  
A:Accession: A23513  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-183 <ARG>  
A:Cross-references: GB:X03591; NID:946740; PIDN:CAA27265.1; PID:946741  
R:Gilln, G.; Bayer, E.A.; Wilchek, M.  
Biochem. J. 269, 527-530, 1990  
A:Title: Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine res  
A:Reference number: S11540; MID:90351377  
A:Accession: S11540  
A:Molecule type: protein  
A:Residues: 54-66 'X', 68 <GIT>  
C:Superfamily: streptavidin  
C:Keywords: biotin; homotrimer  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-183/Product: streptavidin #status predicted <MAT>  
F:67/Binding site: biotin (Tyr) #status predicted  
F:78/Binding site: biotin (Tyr) #status predicted

Query Match 100.0%; Score 936; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 8.7e-67;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVVAIAIVSLTIVSITASADPSKDSKAQVSAEAGITGTWVQLGSTFLVTAAD 60  
|||||  
DB 1 MKRIVVAIAIVSLTIVSITASADPSKDSKAQVSAEAGITGTWVQLGSTFLVTAAD 60  
QY 61 GALTGTESAVGNESRYVTGRDSPAATDGSCTALGWTAVAKNNYRNHSAATWSGGY 120  
|||||  
DB 61 GALTGTESAVGNESRYVTGRDSPAATDGSCTALGWTAVAKNNYRNHSAATWSGGY 120  
QY 121 VGGAERINTOWLLTSGTTEANMKSTLVGHDFPTTKVPSAASIDAACKAGVNNGNPLDA 180  
|||||  
DB 121 VGGAERINTOWLLTSGTTEANMKSTLVGHDFPTTKVPSAASIDAACKAGVNNGNPLDA 180  
QY 181 VQO 183  
|||  
DB 181 VQO 183  
RESULT 2

S57284 Streptavidin v1 precursor - Streptomyces venezuelae  
C:Species: Streptomyces venezuelae  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 10-Dec-1999  
C:Accession: S57284  
R:Bayer, E.A., Kulik, T., Adar, R., Wilchek, M.  
Biochim. Biophys. Acta 1263, 60-66, 1995  
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptomyces  
A:Reference number: S57284; MUID:55359204  
A:Accession: S57284  
A:Molecule type: DNA  
A:Residues: 1-183 <BAV>  
A:Experimental source: strain Tue2460  
C:Superfamily: streptavidin  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-183/Product: streptavidin v1 #status predicted <MAT>

Query Match	99.6%	Score 932;	DB 2;	Length 183;
Best Local Similarity	99.5%	Pred. No.	1.8e-66;	
Matches 182;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

QY	1	MRIIVAAIAVSLTSTYSTASASADPSKDSKAOYSAAEAGITGWNOIGSTFTYVACAD	60	
		1	MRIIVAAIAVSLTSTYSTASASADPSKDSKAOYSAAEAGITGWNOIGSTFTYVACAD	60
QY	61	GALTGYESAVGAEBSRYVLTRGRYDAPATGSGTALGWTYAKNNYRNAHSATITMSGOY	12	
		61	GALTGYESAVGAEBSRYVLTRGRYDAPATGSGTALGWTYAKNNYRNAHSATITMSGOY	12
QY	61	GALTGYESAVGAEBSRYVLTRGRYDAPATGSGTALGWTYAKNNYRNAHSATITMSGOY	12	
		61	GALTGYESAVGAEBSRYVLTRGRYDAPATGSGTALGWTYAKNNYRNAHSATITMSGOY	12
QY	121	VGGAEARINTOMLLTSGTTEANAMKSTLYGHDTFTTKVPSAASIDAAKAGVNGNPDLA	18	
		121	VGGAEARINTOMLLTSGTTEANAMKSTLYGHDTFTTKVPSAASIDAAKAGVNGNPDLA	18
QY	121	VGGTTEARINTOMLLTSGTTEANAMKSTLYGHDTFTTKVPSAASIDAAKAGVNGNPDLA	18	
		121	VGGTTEARINTOMLLTSGTTEANAMKSTLYGHDTFTTKVPSAASIDAAKAGVNGNPDLA	18

27	101	102	103
	111		
Db	181	V00	183

RESULT 3  
S57285  
streptavidin v2 precursor - Streptomyces venezuelae  
C:Species: Streptomyces venezuelae  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 10-Dec-1999  
C:Accession: S57285  
R:Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.  
Biochim. Biophys. Acta 1263, 60-66, 1995  
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom  
A:Reference number: S57284; MUID:9535204  
A:Accession: S57285  
A:Molecule type: DNA  
A:Residues: 1-183 <BAV>  
A:Experimental source: strain Tue2605  
C:Superfamily: streptavidin  
F:1-24/Domain: signal sequence #stratus predicted <SIG>  
;25-183/Product: streptavidin v2 #status predicted <MAT>

Query Match 95.8%; Score 897; DB 2; Length 183;  
 Best Local Similarity 95.1%;  
 Matches 174; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY	1	MRIYVAALAVSLTVSTVSTASADSPKSDSAQVSAAEATITGTWYNOLGSRFEIYTAGD	60
Db	1	MRIYVAALAVSLTVSTVSTASADSPKSDSAQVSAAEATITGTWYNOLGSRFEIYTAGD	60
QY	1	MRIYVAALAVSLTVSTVSTASADSPKSDSAQVSAAEATITGTWYNOLGSRFEIYTAGD	60
Db	1	MRIYVAALAVSLTVSTVSTASADSPKSDSAQVSAAEATITGTWYNOLGSRFEIYTAGD	60
QY	61	GALTGTYSANVGNASRYLVGRDYSAPADDSGTLGTVYANKNYRAHSATTTWSGOY	120
Db	61	GALTGTYSANVGNASRYLVGRDYSAPADDSGTLGTVYANKNYRAHSATTTWSGOY	120
QY	61	GALTGTYSANVGNASRYLVGRDYSAPADDSGTLGTVYANKNYRAHSATTTWSGOY	120
Db	61	GALTGTYSANVGNASRYLVGRDYSAPADDSGTLGTVYANKNYRAHSATTTWSGOY	120
QY	121	VGGAEARINFTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKRAVGNNGPLDA	180
Db	121	VGGAEARINFTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKRAVGNNGPLDA	180
QY	121	VGGAEARINFTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKRAVGNNGPLDA	180
Db	121	VGGAEARINFTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKRAVGNNGPLDA	180

QY	181	VQQ	183
Db	181	VQQ	183

RESULT 4  
S42204

avidin-related protein 4/5 precursor - chicken  
C;Species: Gallus gallus (chicken)

C;date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 07-May-1999  
C;Accession: S42204; S42205

R. Keinänen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Helenius, J. *J. Biochem.* 220, 615-621, 1994

A1:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-  
A1:Reference number: S42201, MUID:94170814

A;Accession: S42204  
A;Molecule type: DNA

A;Residues: 1-150 <KEI>  
A;Cross-references: EMB

A;Experimental source: strain White Leghorn; tissue oviduct  
A;Genetics: CH1

A;Accession: S42205  
A;Molecule type: DN

A;Residues: 1-150 <KEW>  
A;Cross-references: EMBL:Z22882; NID:g311812

A;Experimental source: strain White Leghorn; tissue oviduct  
A;Genetics: CH2

C;Genetics: <CH1>  
A;Gene: avr4

A; Introns: 27/3; 96/1; 136/2  
C; Genetics: <CH2>

A;Gene: avr5  
A;Introns: 27/3; 96/1; 136/2

C;Superfamily:	avidin
F:1-24/Domain:	signal sequence #status predicted <SIG>

F:25-150/Product: avidin-related protein 4/5 #status predicted <MAT>  
F:28-105/Disulfide bonds: #status predicted

100

Query match	15.7%	Score 140.5	DB 2	Length	150
Best Local Similarity	33.8%	Pred. No. 7.5e-05			
Matches 45; Conservative	22;	Mismatches 47;	Indels 19;	Gaps	8;

34 VSAEAGITGTMYNQLGSTFIIVA-GADGALTGTYESAV----GNAESRYVLTRGDSAP 88

```
Db 23 LSARKSLTGWNTNNLGSIMTIRAVNSRGEFTGYLFAVDNPGNITLSPPLGIGHRA - 81
```

89 ATDGGTALGNTVAMKNNYRNAHSATTTMSGQY---GGAEARINTQWLLTSGTTEAN-AM 144

Db 82 ----SQPTFGTVLHW--NF--SESTTVFTGQCFIDRNGKEV-LKTMILLRSSVNDISYDW 133

QY 145 KSTLVGHDFTKV 157

Db 133 KATRVGYNNETRL 145

10

RESULT 3  
VICH

```

avain precursor [validated] - chicken
C:Species: Gallus gallus (chicken)

```

C/Accession: A54975; A27518; A92093; A92092; A03160  
C/Date: 24-Apr-1984 #sequence\_revision 04-Nov-1994

R; Wallén, M.O.; Laakkonen, M.O.; Kulomaa, M.S. submitted to GenBank, January 1994

A/Description: sequence of the chicken egg-white avidin gene.

A;Accession: A549/5  
A;Molecule type: DN/

A/residues: 1-21, S, 23-152 <WALL>  
A/Cross-references: GB:L27818; NID:g450255; PID:g451889

A:Note: difference at position 22 may be due to PCR error in gene sequence



R.Gope, M.L.; Keinonen, R.A.; Kristo, P.A.; Connely, O.M.; Beattie, W.G.; Zarucki-Schu  
Nucleic Acids Res. 15, 3595-3606, 1987  
A:Title: Molecular cloning of the chicken avidin cDNA.  
A:Reference number: A27518; MUID:87203384  
A:Accession: A27518  
A:Molecule type: mRNA  
A:Residues: 1-152 <GDP>  
A:Cross-references: GB:X05343; NID:963071; PIDN:CAA28954.1; PID:963072  
R:Delange, R.J.; Huang, T.S.  
J. Biol. Chem. 246, 686-709, 1971  
A:Title: Egg white avidin. III. Sequence of the 78-residue middle cyanogen bromide pep  
A:Reference number: A92093; MUID:71107558  
A:Accession: A92093  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 25-57, 'T', 59-76, 'E', 78-152 <DEL>  
A:Experimental source: egg white  
A:Note: Approximately 50% of the chains have 58-ile  
R:Huang, T.S.; Delange, R.J.  
J. Biol. Chem. 246, 686-697, 1971  
A:Title: Egg white avidin. II. Isolation, composition, and amino acid sequences of the  
A:Reference number: A92092; MUID:71107557  
A:Accession: A92092  
A:Status: sequences of tryptic peptides  
A:Molecule type: protein  
A:Residues: 25-57, 'T', 59-76, 'E', 78-152 <DEL>  
R:Livnah, O.; Sussman, J.  
Submitted to the Brookhaven Protein Data Bank, April 1993  
A:Reference number: A51448; PDB:AVT  
A:Contents: annotation; X-ray crystallography, 3.0 angstroms, with biotin, residues 27-5  
R:Livnah, O.; Bayer, E.A.; Wilchek, M.; Sussman, J.L.  
Proc. Natl. Acad. Sci. U.S.A. 90, 5076-5080, 1993  
A:Title: Three-dimensional structures of avidin and the avidin-biotin complex.  
A:Reference number: A47554; MUID:93281699  
A:Contents: annotation; X-ray crystallography, 3.0 angstroms  
R:Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M.  
Submitted to the Brookhaven Protein Data Bank, March 1993  
A:Reference number: A51622; PDB:JAVD  
A:Contents: annotation; X-ray crystallography, 2.7 angstroms, with biotin, residues 27-5  
R:Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M.  
Submitted to the Brookhaven Protein Data Bank, March 1993  
A:Reference number: A51623; PDB:JAVE  
A:Contents: annotation; X-ray crystallography, 2.8 angstroms, without biotin, residues 2  
J. Mol. Biol. 231, 698-710, 1993  
A:Title: Three-dimensional structure of the tetragonal crystal form of egg-white avidin  
A:Reference number: A54974; MUID:93294833  
A:Contents: annotation; X-ray crystallography, 2.7 angstroms  
C:Genetics:  
A:Introns: 27/3; 98/1; 138/2  
C:Superfamily: avidin  
C:Keywords: glycoprotein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-152/Product: avidin #status experimental <MAT>  
F:28-107/Disulfide bonds: #status experimental  
F:41/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 15.3%; Score 143.5; DB 1; Length 152;  
Best Local Similarity 33.3%; Pred. No. 0.00013;  
Matches 44; Conservative 19; Mismatches 56; Indels 13; Gaps 7;

OY 34 VSAAEAGICTGTWNOGSGFTIYVAG--GADGALITYESANGNAESRYVLICRYSAPRTDQ 92  
DB 23 LSRKCSLTGKWNDS--IMTIGAVNDGEGFTIYVADNPNGITRPSPLGIQHKR 80  
R:Keinonen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Heleni  
Biochim. Biophys. Acta 1216, 441-445, 1993  
A:Title: Induction of chicken avidin and related mRNAs after bacterial infection.  
A:Reference number: S39799; MUID:94092737  
A:Accession: S39800  
A:Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 71-150 <KUN>  
A:Cross-references: EMBL:221536; NID:965429  
C:Genetics:  
A:Gene: avr3  
A:Introns: 27/3; 96/1; 136/2  
C:Superfamily: avidin

RESULT 6  
S42201  
avidin-related protein 1 precursor - chicken  
N:Alternate names: avr1  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 07-May-1999  
A:Accession: S42201  
R:Keinonen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Heleni  
Eur. J. Biochem. 220, 615-621, 1994  
A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-  
A:Reference number: S42201; MUID:94170814  
A:Accession: S42201  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-150 <KEI>  
A:Cross-references: EMBL:221611  
C:Genetics:  
A:Introns: 27/3; 96/1; 136/2  
C:Superfamily: avidin  
C:Keywords: glycoprotein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-150/Product: avidin-related protein 1 #status predicted <MAT>  
F:28-105/Disulfide bonds: #status predicted  
F:54, 67, 93/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.0%; Score 121.5; DB 2; Length 150;  
Best Local Similarity 32.8%; Pred. No. 0.0069;  
Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;

OY 34 VSAAEAGICTGTWNOGSGFTIYVAG--DGALITGESAV---GNAESRYVLITGYDS 86  
DB 23 LSRKCSLTGKWNDS--IMTIGAVNDGEGFTIYVADNPNGITRPSPLGIQHKR 80  
OY 87 APATDSCGTALGWTVAKNRYRNAHSATWSGO--YV--GGAERINTQWLITSGTTE--AN 142  
DB 81 A-----CQPFPGFTVHK--NF--SESTSVVGCFVDSKSEV-LTKRWLQRLAVNDISD 130  
OY 143 AMKSTLVGHDTFVK 156  
DB 131 DMKATRGVNDFFR 144

RESULT 7  
S42203  
avidin-related protein 3 precursor - chicken  
N:Alternate names: avr3 protein  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 07-May-1999  
A:Accession: S42203; S39800  
R:Keinonen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Heleni  
Eur. J. Biochem. 220, 615-621, 1994  
A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-  
A:Reference number: S42201; MUID:94170814  
A:Accession: S42203  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-150 <KEI>  
A:Cross-references: EMBL:221612; NID:965432  
R:Kunns, T.A.; Wallen, M.J.; Kuloma, M.S.  
Biochim. Biophys. Acta 1216, 441-445, 1993  
A:Title: Induction of chicken avidin and related mRNAs after bacterial infection.  
A:Reference number: S39799; MUID:94092737  
A:Accession: S39800  
A:Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 71-150 <KUN>  
A:Cross-references: EMBL:221536; NID:965429  
C:Genetics:  
A:Gene: avr3  
A:Introns: 27/3; 96/1; 136/2  
C:Superfamily: avidin





A:Accession: B43855  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1477 <BAR>  
 A>Note: sequence inconsistent with the nucleotide translation  
 A>Note: sequence extracted from NCBI backbone (NCBIN:89237, NCBI:89240)

Query Match 10.6%; Score 99.5; DB 2; Length 1477;  
 Best Local Similarity 26.2%; Pred. No. 4.1;  
 Matches 50; Conservative 26; Mismatches 88; Indels 27; Gaps 6;

QY 11 VSLTTSITASASADPSKSKAQSAAEAGI---TGTWYNOL-GSTFIYTAGDGLTGT 66  
 Db 1177 ISGNVSVSATVLTITKSGSKIEAKSGEANTYSATGTTGCTISGNTVATANA-GDLTIG 1235  
 QY 67 YESANGNAESRYVLTGRYDSAPATDGS-----TALGWTVAAMKNYRNAHSATTW 116  
 Db 1236 NGAELNATEGAATTATGCTGNTLTTEAGSSITSTKGVLDLLAONGSIAGSINAAVTLMTTG 1295  
 QY 117 SCQYVGAEARINTQWLITSGTTEANAMKSTLVGHDFETTKKPSA-----ASIDAKKA 170  
 Db 1296 TLTYTAGSDIR-----ATSGTLVINAKDAKLNGDASGDSSTEVANVAGSGSVTAATSS 1349  
 QY 171 GYNNGNPLDAY 181  
 Db 1350 SVNITGDLNTV 1360

RESULT 14  
 A:Accession: B43412  
 A:hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
 C:Accession: A83412  
 R:Stover, C. K.; Pham, X. Q.; Erwin, A. L.; Mizoguchi, S. D.; Warren, P.; Hickey, M. J.; B.  
 adman, S.; Yuan, Y.; Brody, L. L.; Coulter, S. N.; Folger, K. R.; Kas, A.; Lardig, K.; Lam,  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
 A:Reference number: A82950  
 A:Accession: A83412  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2468 <STC>  
 A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AA05263.1; GSPDB:GN001  
 A:Experimental source: strain PAO1  
 C:Genetics:  
 A:Gene: PA1874

Query Match 10.6%; Score 99.5; DB 2; Length 2468;  
 Best Local Similarity 24.9%; Pred. No. 6.9;  
 Matches 61; Conservative 19; Mismatches 70; Indels 95; Gaps 12;

QY 14 TTVSITASASA-----DPSKDSKAQVSAEAG-ITGTWYNOLGSTFIYTAG--- 58  
 Db 1381 TVVNVITASDAAGNTSLPATTVYDSSLPISIPQVPSNGSVISGT--ADAGNTTIIITDGN 1438  
 QY 59 -----ADGA-----LTGTESAAGNAESRYVLTGRYDSAPA----- 89  
 Db 1439 PIGQVTAADSSGNMSTTPGJPLPDGTIVNVNARSPS-----NVDSPAVITVDGVAPAAP 1492  
 QY 90 -----TDGSGTA-LGWTVAAMKNYRNAHSATTWGSQYVGAEARINTQWLITSGTTEA 141  
 Db 1493 VIDPSNGTISGTAEGATVI-----LTDGGGNPIGQATADGSGNMWTFPTSTPLA 1542  
 QY 142 NAMKSTLVGHDFETTKV-PSAASIDA-----AKKAGV-----NNGN 176  
 Db 1543 NCTVINAAYADPAGMTSGPASTVVDALAPAPVNPNSNGVVISGTAEGATVILIDGN 1602  
 QY 177 PLDAY 181

Db 1603 PIGOV 1607

RESULT 15  
 F70722

probable glcB protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: F70722  
 R:Cooper, R.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 Rastndream, M. A.; Rogers, J.; Ruter, S.; Seeger, K.; Skellon, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: F70722  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-741 <COL>  
 A:Cross-references: GB:Z78020; GB:AL123456; NID:g3261625; PIDN:CAB01465.1; PID:e25767  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: glcB

Query Match 10.3%; Score 96.5; DB 2; Length 741;  
 Best Local Similarity 26.5%; Pred. No. 3.4;  
 Matches 56; Conservative 28; Mismatches 86; Indels 41; Gaps 11;

QY 9 INVSITTSITASASAD---PSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGDGLTGT 65  
 Db 100 ITTSGVDAETTTTAGPOLVVPVPLNARFALNANAR-WGSLYDALYGTDIVPE-TDGAERG 157  
 QY 66 -TYESAAGNAESRYVLTGRYDSAPATDGS-GTALGWT-----VAMKNYRNAHSATTW 116  
 Db 158 PTYNKVRGKVIAYAKRFLDDSVPLSSGSFGDATGTGYDDGQVVALPKSTGLANPGOF 217  
 QY 117 SCQYVGAEARINTQWLITS-----GTTEANAMKSTLV--GHDFTTKVXP 159  
 Db 218 AG-YTGAAES--PTSVLLINHGILHIEILLIDPESQVGTTRACGVKQVILLESATTTIMDFED 274  
 QY 160 SAASIDAKKA-----GYNNGNPLDAYQO 183  
 Db 275 SYAAVDAAADKVLGYRNWGLNKGDLAAAYDK 305

Search completed: April 6, 2001, 00:17:14  
 Job time: 10899 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 00:17:16 ; Search time 67.49 Seconds  
(without alignments)  
87.566 Million cell updates/sec

Title: US-09-589-870-2  
Perfect score: 936  
Sequence: 1 MRKIVAAVAIVSLTTSITA.....IDAKKAGVNNGNPLDAVQ 183

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	936	100.0	183	1 SAV_STRAV	P22629 streptomyc
2	932	99.6	183	1 SAV1_STRVL	Q53533 streptomyc
3	897	95.8	183	1 SAV2_STRVL	Q53533 streptomyc
4	146.5	15.7	150	1 AVR4_CHICK	P56734 gallus gall
5	143.5	15.3	152	1 AVTD_CHICK	P02701 gallus gall
6	124.5	13.3	150	1 AVR6_CHICK	P56735 gallus gall
7	121.5	13.0	150	1 AVR1_CHICK	O13153 gallus gall
8	120.5	12.9	150	1 AVR3_CHICK	P56736 gallus gall
9	120.5	12.9	150	1 AVR3_CHICK	P56733 gallus gall
10	120	12.8	1064	1 FBP1_STRPU	P10079 strongyloc
11	116.5	12.4	150	1 AVR2_CHICK	P56732 gallus gall
12	101.5	10.3	1025	1 SLAP_CAUCR	P35828 caulobacter
13	96.5	10.3	741	1 MAS2_MYCTU	Q50596 mycobacteri
14	93	9.9	1225	1 Y309_MYCGE	P47551 mycoplasma
15	91.5	9.8	1120	1 STRF_ECOLI	P76072 escherichia
16	91.5	9.8	2249	1 190K_RICRI	P15921 rickettsia
17	90	9.6	1150	1 APMU_PIG	P12021 sus scrofa
18	89	9.5	255	1 CBPM_STRAL	P00733 streptomyc
19	89	9.5	289	1 HRPD_CAUCR	Q45573 caulobacter
20	89	9.5	488	1 PHB_ALCFA	P12625 alcalligenes
21	88.5	9.5	465	1 SLAP_LACBR	Q05044 lacobacilli
22	87.5	9.3	457	1 PRTC_SRRGR	P52320 streptomyc
23	87	9.3	666	1 ALYS_ENTHR	P39046 enterococcu
24	86.5	9.2	500	1 ABFL_TTRIR	Q92455 trichoderma
25	86	9.2	598	1 VAUF_SCHPO	Q10168 schizosacch
26	85	9.1	797	1 VGLX_HSEVB	P28968 equine herp
27	85	9.1	872	1 GUNA_CELFI	P50401 cellulomona
28	84.5	9.0	962	1 GUNA_PSEFL	P10476 pseudomona
29	83.5	8.9	489	1 FLIC_SALTY	P06179 salmonella
30	83.5	8.9	500	1 XYU1_TTRKO	P48792 trichoderma
31	83.5	8.9	575	1 FLA2_CAMJE	P22251 campylobact
32	83.5	8.9	575	1 FLB2_CAMJE	P22252 campylobact
33	83.5	8.9	1645	1 OMPB_RICTY	P96989 r outer mem

34	83	8.9	610	1 CHIT_STRPL	P11220 streptomyc
35	82.5	8.8	297	1 PRTA_SRRGR	P00776 streptomyc
36	82.5	8.8	331	1 ADT1_WHEAT	Q41629 triticum ae
37	82.5	8.8	563	1 HEMA_IACKV	P09343 influenza a
38	82.5	8.8	616	1 MURA_STRCM	Q05064 streptomyc
39	82	8.8	511	1 GUNB_PSEFL	P18126 pseudomona
40	82	8.8	570	1 FBP3_STRPU	P49013 strongyloc
41	82	8.8	572	1 FLAB_CAMCO	P27053 campylobact
42	82	8.8	572	1 FLAB_CAMCO	P18245 campylobact
43	82	8.8	704	1 OE66_NPVAC	Q00704 autographa
44	82	8.8	1090	1 GUXB_CELFI	P50899 cellulomona
45	82	8.8	2003	1 YDBA_ECOLI	P33866 escherichia

## ALIGNMENTS

RESULT	1	SAV_STRAV	STANDARD:	PRT:	183 AA.
AC	P22629;	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	STREPTAVIDIN PRECURSOR.				
OS	Streptomyces avidinii.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 24-64.				
RP	MEDLINE-86148514; PubMed-3951999;				
RX	Argarane C.E., Kuntz J.D., Birken S., Axel R., Cantor C.R.;				
RT	"Molecular cloning and nucleotide sequence of the streptavidin gene.";				
RL	Nucleic Acids Res. 14:1871-1882(1986).				
RN	[2]				
RP	IMPORTANCE OF TRP IN BIOTIN-BINDING.				
RX	MEDLINE-89134083; PubMed-3223904;				
RT	Gitlin G., Bayer E.A., Wilchek M.;				
RT	"Studies on the biotin-binding site of streptavidin. Tryptophan				
RT	residues involved in the active site.";				
RL	Biochem. J. 256:279-282(1988).				
RN	[3]				
RP	IMPORTANCE OF TYR IN BIOTIN-BINDING.				
RX	MEDLINE-90351377; PubMed-2386489;				
RT	Gitlin G., Bayer E.A., Wilchek M.;				
RT	"Studies on the biotin-binding sites of avidin and streptavidin.				
RT	Tyrosine residues are involved in the binding site.";				
RL	Biochem. J. 269:527-530(1990).				
RN	[4]				
RP	CELL-BINDING.				
RX	MEDLINE-90358825; PubMed-2390089;				
RT	Alon R., Bayer E.A., Wilchek M.;				
RT	"Streptavidin contains an RYD sequence which mimics the RGD receptor				
RT	domain of fibronectin.";				
RL	Biochem. Biophys. Res. Commun. 170:1236-1241(1990).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY OF 37-157.				
RX	MEDLINE-89100243; PubMed-2911722;				
RT	Weber P.C., Ohlendorf D.H., Wendoloski J.J., Salem F.R.;				
RT	"Structural origins of high-affinity biotin binding to streptavidin.";				
RL	Science 243:85-88(1989).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.				
RX	MEDLINE-97337436; PubMed-914176;				
RT	Freitag S., Je T'rong I., Klumb L., Stayton P.S., Stenkamp R.E.;				
RT	"Structural studies of the streptavidin binding loop.";				
RL	Protein Sci. 6:1157-1166(1997).				
RN	[7]				
RP	X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS) OF 37-157.				
RX	MEDLINE-97294734; PubMed-9148939;				
RT	Katz B.A., Cass R.J.;				
RT	"In crystals of complexes of streptavidin with peptide ligands				
RT	containing the HPQ sequence the pKa of the peptide histidine is less				





QY	1	MKKIYVAIAIVSLTTVTSTTASADPSKDSKAQVSAAEAGITGTWYNQLGSTFIYTAGAD	60
Db	1	MKKIYVAIAIVSLTTVTSTTASADPSKDSKAQVSAAEAGITGTWYNQLGSTFIYTAGAD	60
QY	61	GALGTCTSSAVGNMESRVLVTCGRYDSAPATGSGGALGWTVAAMKNNYNNAHSATTWMSQY	120
Db	61	GALGTCTSSAVGNMESRVLVTCGRYDSAPATGSGGALGWTVAAMKNNYNNAHSATTWMSQY	120
QY	121	VCGAEARINTQWLITSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAARKAGNNGNPIDA	180
Db	121	VCGTARARINTQWLITSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAARKAGNNGNPIDA	180
QY	181	VQO 183	
Db	181	VQO 183	
RESULT	3		
SAV2_STRVL	SAV2_STRVL	STANDARD:	PRT: 183 AA.
AC	053533:		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	STREPTAVIDIN V2 PRECURSOR (SA V2).		
OS	Streptomyces violaceus (Streptomyces venezuelae).		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95359204; PubMed=7632734;		
RA	Bayer E.A., Kulik T., Adar R., Mlilchek M.;		
RT	"Close similarity among streptavidin-like, biotin-binding proteins from Streptomyces."		
RL	Biochim. Biophys. Acta 1263:60-66(1995).		
CC	-1- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.		
CC	FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE		
CC	MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).		
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: SECRETED.		
CC	-1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; S78782; AAB35016.1; -.		
DR	HSSP; P22629; 1PTS.		
DR	INTERPRO: IPR000089; -.		
DR	PFAM; PF01382; Avidin; 1.		
DR	PROSITE: PS00577; AVIDIN; 1.		
KW	Signal; Biotin.		
FT	SIGNAL	1 24	BY SIMILARITY.
FT	CHAIN	25 183	STREPTAVIDIN V2.
FT	BINDING	67 67	INVOLVED IN BIOTIN BINDING (BY
FT	BINDING		SIMILARITY).
FT	BINDING	78 78	INVOLVED IN BIOTIN BINDING (BY
FT	BINDING		SIMILARITY).
FT	BINDING	116 116	INVOLVED IN BIOTIN BINDING (BY
FT	BINDING		SIMILARITY).
FT	BINDING	132 132	INVOLVED IN BIOTIN BINDING (BY
FT	BINDING		SIMILARITY).
FT	BINDING	144 144	INVOLVED IN BIOTIN BINDING (BY
FT	BINDING		SIMILARITY).
QO	SEQUENCE	183 AA; 18833 MW; FEAFEPFEAE4ECCA CRC64;	

Query Match 95.8%; Score 897; DB 1; Length 183;  
 Best Local Similarity 95.1%; Pred. No. 3,9e-66;  
 Matches 174; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRKIVAAIVSLTFTVITASADSPKSKAQAASAAGITGTWYNQSGTFTVTAGAD 60  
 DB 1 MRKIVAAIVSLTFTVITASADSPKSKAQAASAAGITGTWYNQSGTFTVTANAD 60  
 QY 61 GALTGTESAVGNAESRYVLTGRDSPAATDGSCTALGVTYAKNNRNMAHSATTGSGQY 120  
 DB 61 GSLGTYESAVGNAESRYVLTGRDSPAATDGSCTALGVTYAKNNRNMAHSATTGSGQY 120  
 QY 121 VGGAEARINTOWILTSGTTEANAKSTLVGHDTFTKYKPSAASIDAAKKGAVNNGNPLDA 180  
 DB 121 VASGEARINTOWILTSGTTEANAKSTLVGHDTFTKYKPSAASIDAAKKGAVNNGNPLDA 180  
 QY 181 VQQ 183  
 DB 181 VQQ 183

RESULT 4  
 AV4\_CHICK STANDARD; PRT; 150 AA.

AC P56734;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE AVIDIN-RELATED PROTEIN 4/5 PRECURSOR.  
 GN AV4 AND AV4S.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WHITE LEHORN; TISSUE-OVIDUCT;  
 RX MEDLINE=94170814; PubMed=8125122;  
 RA Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,  
 RA Toimela T.A., Helenius M.A., Kulomaa M.S.;  
 RT "Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-5."  
 RL Eur. J. Biochem. 220:615-621(1994).  
 CC -1- MISCELLANEOUS: THE SEQUENCES OF THE CODING REGIONS OF GENES AV4 AND AV4S ARE IDENTICAL.  
 CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: Z22883; -; NOT ANNOTATED\_CDS.  
 DR INTERPRO: IPR000088; -;  
 DR PRINTS: PR00709; AVIDIN.  
 DR PROSITE: PS00577; AVIDIN, FALSE\_NEG.  
 KW Biotin; Signal; Multigene family.  
 FT CHAIN 1 24 POTENTIAL.  
 FT SIGNAL 25 150 AVIDIN-RELATED PROTEIN 4/5.  
 FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 150 AA; 16644 MW; 9AC6C6310EFE13A CRC64;

Query Match 15.7%; Score 146.5; DB 1; Length 150;  
 Best Local Similarity 33.8%; Pred. No. 1.8e-05;  
 Matches 45; Conservative 22; Mismatches 47; Indels 19; Gaps 8;

QY 34 VSAEAGITGTWYNQSGTFTVTA-GADGALTGTYESA-----GNAESRYVLTGRDSPA 88  
 DB 23 LSARKCSLTGKWNINNGISMTIRAVNSRGFTGTLYLAVDNPGNTTSLPLGIQHKRA- 81  
 QY 89 ATDGSCTALGVTYAKNNRNMAHSATTWSGQY---GGAEARINTOWILTSGTTEAN-AW 144  
 DB 82 ----SQTFEFTYHW--NF--SESTVFTGQCFIDRNGKEV-LKTMWLLNSSVNDISYDM 132  
 QY 145 KSTLVGHDTFTKY 157  
 DB 133 KATRVGYNNFRL 145

RESULT 5  
 AVID\_CHICK STANDARD; PRT; 152 AA.

AC P02701; Q91958;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE AVIDIN PRECURSOR.  
 GN AV4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=87203384; PubMed=3575102;  
 RX Gope M.L., Keinaenen R.A., Kristo P.A., Connely O.M., Beattie W.G.,  
 RA Zarnucki-Schulz T., O'Malley B.W., Kulomaa M.S.;  
 RT "Molecular cloning of the chicken avidin cDNA."  
 RL Nucleic Acids Res. 15:3595-3606(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90355928; PubMed=2143802;  
 RX Chandra G., Gray J.G.;  
 RT "Cloning and expression of avidin in Escherichia coli."  
 RL Meth. Enzymol. 184:70-79(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WHITE LEHORN; TISSUE-OVIDUCT;  
 RX MEDLINE=95394357; PubMed=7665080;  
 RA Wallen M.J., Laukkanen M.O., Kulomaa M.S.;  
 RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene and its relationship with the avidin-related genes Av1-Av5."  
 RL Gene 161:205-209(1995).  
 RN [4]  
 RP SEQUENCE OF 25-152.  
 RX MEDLINE=71107558; PubMed=5100763;  
 RA Delange R.J., Huang T.-S.;  
 RT "Egg white avidin. 3. Sequence of the 78-residue middle cyanogen bromide peptide. Complete amino acid sequence of the protein subunit."  
 RL J. Biol. Chem. 246:698-709(1971).  
 RN [5]  
 RP IMPORTANCE OF TYR IN BIOTIN-BINDING.  
 RX MEDLINE=90351377; PubMed=2386489;  
 RA Gilin G., Bayer E.A., Wilchek M.;  
 RT "Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine residues are involved in the binding site."  
 RL Biochem. J. 269:527-530(1990).  
 RN [6]  
 RP BIOTIN-BINDING STUDIES.  
 RX MEDLINE=91378911; PubMed=1898347;  
 RA Hiller Y., Bayer E.A., Wilchek M.;  
 RT "Studies on the biotin-binding site of avidin. Minimized fragments that bind biotin."  
 RL Biochem. J. 278:573-585(1991).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=93281699; PubMed=8506353;  
 RA Livnah O., Bayer E.A., Wilchek M., Sussman J.L.;

RT	"Three-dimensional structures of avidin and the avidin-biotin complex.";
RL	Proc. Natl. Acad. Sci. U.S.A. 90:5076-5080(1993).
RN	[8]
RP	X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX	MEDLINE=9329483; PubMed=8515446;
RA	Pugliese L., Coda A., Malcovati M., Bolognesi M.;
RA	Siccardi A.G., Pagnelli G., Losso R., Apreda B., Bolognesi M.,
RA	Sidoli A., Arasio P.;
RT	"Biochemical characterization and crystal structure of a recombinant hen avidin and its acidic mutant expressed in Escherichia coli.";
RL	Eur. J. Biochem. 256:453-460(1996).
-1	FUNCTION: THE BIOLOGICAL FUNCTION OF AVIDIN IS NOT KNOWN. FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE MOLECULE OF BIOTIN PER SUBUNIT OF AVIDIN).
-1	TISSUE SPECIFICITY: SYNTHESIZED IN HEN OVIDUCT AND CONCENTRATED IN EGG WHITE (WHERE IT REPRESENT 0.05% OF THE TOTAL PROTEIN).
-1	SIMILARITY: BELONGS TO THE AVIDIN/SREPTAVIDIN FAMILY.
-1	DATABASE: NME-Worthington enzyme manual;
CC	WWW-"http://www.worthington-biochem.com/manual/A/Av.html".
CC	-1 DATABSE: NME-Prozyme technical fact sheet;
CC	WWW-"http://www.prozyme.com/technical/avidate.html".
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X05343: CA28954.1; -.
DR	EMBL; L27818: AMB5973.1; -.
DR	PIR; A03180; VICH.
DR	PIR; A27518; A27518.
DR	PIR; S11540; S11540.
DR	PDB; 2AVI; 15-JUL-93.
DR	PDB; 1AVD; 31-JAN-94.
DR	PDB; 1AVE; 31-JUN-94.
DR	PDB; 1RAW; 15-JUL-98.
DR	PDB; 2CAM; 15-JUL-98.
DR	INTERPRO: IPR000088; .
DR	PFAM: PF01382; Avidin; 1.
DR	PRINTS: PR00709; AVIDIN.
DR	PROSITE: PS00577; AVIDIN; 1.
KW	Glycoprotein; Signal; Biotin; 3D-structure.
FT	SIGNAL 1 24
FT	CHAIN 25 152
FT	DISULFID 28 107
FT	CARBOHYD 41 41
FT	BINDING 57 57
FT	VARIANT 58 58
FT	CONFLICT 22 22
FT	CONFLICT 77 77
FT	STRAND 32 36
FT	TURN 37 38
FT	STRAND 41 44
FT	TURN 49 50
FT	STRAND 54 58
FT	STRAND 71 74
FT	STRAND 77 77
FT	TURN 83 84
FT	STRAND 87 93
FT	STRAND 100 109
FT	N-LINKED (GLUCNA...)
FT	I-> T (IN APPROX. 50% OF THE CHAINS).
FT	G-> S (IN REF. 3).
FT	E-> Q (IN REF. 2 AND 3).

FT	STRAND	115	124	
FT	HELIX	130	135	
FT	STRAND	137	146	
SO	SEQUENCE	152 AA;	16769 MM;	1D5A4491D5EPD5C CRC64;

Query Match	15.3%	Score 143.5	DB 1	Length 152
Best Local Similarity	33.3%	Pred. No. 32e-05		
Matches 44	Conservative 19	Mismatches 56	Indels 13	Gaps 7

QY VSAAEAGLIGTWINDUSTIFLIA-GAAGALIGITSAVGAESKRIVLGRIUSAPADUG 92  
:: : :: | | | | : | | | | :  
Db 23 LSAKRCSLGHWINDGSNMFTGAVNSGEFTGYITAV-TATSNEIKESP LHGTENTIN 81

Db 82 KRTPTEGFTVNWK----FSESTTVFTGOCFIDRNGKEV-LKTMMLLRSSVNDIGDDMKA 130

DB 137 TRVGNIETRLR 148

RESUL1	6		
AVR6_CHICK			
ID	AVR6_CHICK	STANDARD;	PRT; 150 AA.
AC	P56735;		

30-MAY-2000 (Rel. 39, Last sequence update)  
01-OCT-2000 (Rel. 40, Last annotation update)  
DE AVILIN-RELATED PROTEIN 6 PRECURSOR.

05 Gallus gallus (Chicken).  
0C Euryarchaea; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
0C Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae  
0C Gallus

KN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RHODE ISLAND;  
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R

RT "Characterization and chromosomal localization of the chicken avidin  
RT gene family.";  
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RL

CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstac  
CC the European Bioinformatics Institute. There are no restrictions on  
CC

CC use by non-profit institutions as long as its content is in the  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ237658; CAB39893.1; -.  
DR INTERPRO; IPR00008; -.  
DR PRINTS; PRO0709; AVIDIN.  
SOURCE: 000077 AVIDIN; 1

Accession	Protein	Length	Source
KW	Biotin; signal; Multigene family.		
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	25	AVIDIN-RELATED PROTEIN 6.
FT	CARBOHYD	54	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	SEQUENCE	150 AA:	16528 MW:	COF3357D6291BI50 CRC64:	N-LINKED (GLCNAC. . . ) (POTENTIAL).
1	CHANDID	20			
2	CARBOHYD	141			
3	SEQUENCE	150 AA:	16528 MW:	COF3357D6291BI50 CRC64:	N-LINKED (GLCNAC. . . ) (POTENTIAL).

Query Match	13.3%	Score 124.5;	DB 1;	Length 150;
Best Local Similarity	32.8%	Pred. No. 0.0011;		
Matches 44;	Conservative 25;	Mismatches 42;	Indels 23;	Gaps

```

34 VSAAEGLIGIMINLGGIFIVTGA--DGALIGIIESAV---GNAESKIVLIGRID 88
      :|| : ||| | ||| | :|| || : ||| :|| || : || :
09

```

```

Db      23 LSARCSLTGEMDNLNGS--IMTIGAVNDNGEENGNYITAAVDNPNENIKSLPGLGIQHR 80
OY      87 AATDSSGALGTAYAKNNRYNAHSATTWSG-YV--GGAERINTOMLLISGTTT-AN 142
DB      81 A-----COPFTFFTYHW--NF--SESTSVFVGCCFYDRSGKEY-LKTKWLORLAVDISD 130
OY      143 AMKSTLVGHDTFTK 156
DB      131 DMKATPRVGYNNFTFR 144

RESULT 7
AVR1_CHICK STANDARD; PRT; 150 AA.
AC 013153:
DT 15-JUL-1999 (Rel. 38, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AVIDIN-RELATED PROTEIN 1 PRECURSOR.
GN AVR1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVIDUCT;
RX MEDLINE=86260103; PubMed=28386690;
RA Keinaenen R.A., Laukkanen M.-L., Kulmaa M.S.,
RA Toimela T.A., Helenius M.A., Kulmaa M.S.;
RT "Molecular cloning of three structurally related genes for chicken
   avidin."
RL J. steroid Biochem. 30:17-21(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEHORN; TISSUE=OVIDUCT;
RX MEDLINE=94170814; PubMed=8125122;
RA Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
RA Toimela T.A., Helenius M.A., Kulmaa M.S.;
RT "Molecular cloning and nucleotide sequence of chicken avidin-related
   genes 1-5."
RL Eur. J. Biochem. 220:615-621(1994).
RN [3]
RC -1 SIMILARITY: BELONGS TO THE AVIDIN-STREPTAVIDIN FAMILY.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z21611; ? NOT ANNOTATED_CDS.
DR EMBL; Z97063; CAB09798.1; -.
DR INTERPRO; IPR000088; -.
DR PRINTS; PR00709; AVIDIN.
DR PROSITE; PS00577; AVIDIN; 1.
DR Biotech; Signal; Multigene family.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 1 150 AVIDIN-RELATED PROTEIN 1.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . . ) (POTENTIAL).
SQ SEQUENCE 150 AA; 16468 MW; 77548380B995B1f6 CRC64;

Query Match 13.0%, Score 121.5; DB 1; Length 150;
Best Local Similarity 32.8%; Pred. No. 0.0019;
Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;
```

[illegible][illegible]

```

ID AVB3_CHICK STANDARD; PRT; 150 AA.
AC P56733;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AVIDIN-RELATED PROTEIN 3 PRECURSOR.
GN AVR3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-OVIDUCT;
RX MEDLINE-94170814; PubMed-8125122;
RA Keiminen R.A., Wallen M.J., Kristo P.A., Laakkonen M.O.,
RA Toimela T.A., Helenius M.A., Kulomaa M.S.;
RT "Molecular cloning and nucleotide sequence of chicken avidin-related
RT genes 1-5."
RL Eur. J. Biochem. 220:615-621(1994).
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 221612; -; NOT_ANNOTATED_CDS.
DR EMBL: 221536; -; NOT_ANNOTATED_CDS.
DR INTERPRO: IPR000088; -.
DR PRINTS: PRO0709; AVIDIN.
DR PROSITE: PS00577; AVIDIN.
DR KMW Biotin; Signal; Multigene family.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 150 AVIDIN-RELATED PROTEIN 3.
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 150 AA; 16657 MW; 84DC03926A6BE21C CRC64;

Query Match 12.9%; Score 120.5; DB 1; Length 150;
Best Local Similarity 30.3%; Pred. No. 0.0023;
Matches 40; Conservative 24; Mismatches 49; Indels 19; Gaps 8;

OY 34 VSAEAGITGTWYNQGLSTFTVTA-GADGALITGYESAV-----GNAESRYVLTGRYSAP 88
DB 23 LSAKRCGLTQKNNNNGSMTIRAVNRRGEFAGTYLTVADVNDGNCNKLSPGLGIQHKRA- 81
OY 89 ATDGSGLALGTVAAKNNYNNAHSATWSGOYV---GGAERINTOMLLTSGTTE-ANAW 144
DB 82 ----CQPTFGFTVHM--NF--SESTSVFVGQCFIDRSKGEV-LTKMVLQRLAVDIDSDW 132
OY 145 KSTLVGHDTFTK 156
DB 133 KATRVGYNNFTR 144

RESULT 10
FBP1_STRPU STANDARD; PRT; 1064 AA.
ID FBP1_STRPU
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FIBROPELIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE (UEGF-1).
GN EGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Echinacea; Echinoidea; Strongylocentrotidae;

```

```

OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90112459; PubMed-2514273;
RA Delgado-Illio-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;
RT "Structural analysis of the UEGF gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats."
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE-87319677; PubMed-3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor."
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE-89196806; PubMed-2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin."
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE-91285254; PubMed-2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo."
RL Dev. Biol. 146:89-99(1991).
CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
CC REPEATS.
CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08692; AAA62164.1; -
DR EMBL: L08692; AAA62163.1; -
DR EMBL: X17530; CAA35571.1; -
DR EMBL: M17421; AAA30050.1; -
DR EMBL: X17533; CAA35573.1; -
DR PIR: A29316; A29316.
DR HSSP: P01132; IEPH.
DR INTERPRO: IPR000088; -.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR000859; -.
DR INTERPRO: IPR001438; -.
DR INTERPRO: IPR001881; -.
DR PFAM: PF01382; Avidin; 1.
DR PFAM: PF00431; Cub; 1.

```

Query Match	12.8%;	Score 120;	DB 1;	Length 1064;
Best Local Similarity	28.4%;	Pred. No. 0.02;		
Matches 40;	Conservative 24;	Mismatches 53;	Indels 24;	Gaps
QY 34 VSAEAG---ITGWNQGLSTFLVAGADGALGTG---ESAVGNASRYL---IGR 83				
Db 931 VNCSEVGFCDLEGKMWINECDQYTTTKTSGMLGDIYMYNEPALCYAPFVVGYSNN 990				
QY 84 YDSAPATDGSGLTALGWTAVANKNRYRNASHATVTSQO-YVGAEARINTOMLTS-GTTEA 141				
Db 991 YDF-----PSGFLVY---RDNGQSTTSMTGCGCHLDDGEVLYTWTIMNVSTIQ 1038				
QY 142 NAMKSTLVGHDFTKVPSAA 162				
Db 1039 DIRKSNMVGODKWTRYEOSIA 1059				
RESULT 11				
ID AVR2-CHICK STANDARD; PRT; 150 AA.				
AC P56732;				
DT 30-MAY-2000 (Rel. 39, Created)				
DT 30-MAY-2000 (Rel. 39, Last sequence update)				
DT 30-MAY-2000 (Rel. 39, Last annotation update)				
DE AVIDIN-RELATED PROTEIN 2 PRECURSOR.				
GN AVR2.				
OS Gallus gallus (Chicken)..				
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC Gallus.				
CC [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN-WHITE LEGHORN; TISSUE=OVIDUCT;				
RX MEDLINE=94170814; PubMed=8125122;				
RA Keinonen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,				
RA Toimela T.A., Helenius M.A., Kulomaa M.S.;				
RT "Molecular cloning and nucleotide sequence of chicken avidin-related				
RT genes 1-5."				
RL Eur. J. Biochem. 220:615-621(1994).				
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY				

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z21554; -; NOT_ANNOTATED_CDS.
DR EMBL: Z21535; -; NOT_ANNOTATED_CDS.
DR INTERPRO: IPR000088; -.
DR PRINTS: PR00709; AVIDIN.
DR PROSITE: PS00577; AVIDIN. FALSE NEG.
KW Biotin; signal; Multigene family.
FT SIGNAL 1 24
FT CHAIN 25 150
FT CARBOHYD 67 67
FT CARBOHYD 93 93
FT CARBOHYD 93 93
FT SEQUENCE 150 AA; 16434 MW; 763D1E2B1A93A66D CRC64;
SQ
Query Match 12.4%; Score 116.5; DB 1; Length 150;
Best Local Similarity 32.8%; Pred. No. 0.0049;
Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;
QY 34 VSAEAGITGTYWYNOLGSGFIYTAGA---DGAITGYESAV---GNASRYVLTGRYDS 86
DB 23 LSNRKSLTGEMNDLGS--IMTIGVNDNGEPDGYITAVADNPNTILSPLLGIQHR 80
QY 87 APATDSSGALGVTVMKNNYRNHNSATTWSCQ-YV--GGAERINTQWLLTSGTTE-AN 142
DB 81 A-----SQPTFGFTVM--NF--SESTSVFVGCFDRSGKEY-LTKMLQLRLAVDISD 130
QY 143 AMKSTLVGHDTFTK 156
DB 131 DWIATRVGNNDFTK 144
RESULT 12
SLAP_CAUOCR STANDARD; PRT; 1025 AA.
AC P35828; O46015;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).
GN R5AA.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE-93007489; PubMed-1393820;
RA Gilchrist A., Fisher J.A., Smit J.K.;
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
RT crescentus paracrystalline surface layer protein.";
RL Can. J. Microbiol. 38:193-202(1992).
RN [2]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE-89008089; PubMed-3049545;
RA Fisher J.A., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
RT Caulobacter crescentus.";
RL J. Bacteriol. 170:4706-4713(1988).
RN [3]
RP CHARACTERIZATION.
RC STRAIN-CB15;
RX MEDLINE-98292737; PubMed-9620954;
RA Arizm P., Smit J.K.;
RT "The Caulobacter crescentus paracrystalline S-layer protein is

```

```

RT secreted by an ABC transporter (type I) secretion apparatus.";
RL J. Bacteriol. 180:3062-3069(1998)
CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF062345; AAC38665.1; -.
DR HSSP: P01549; 2MCM.
DR INTERPRO: IPR001343; -.
DR PFAM: PF00353; hemolysinCbind; 1.
DR PRINTS: PR00313; CABNDNGRPT.
KW Cell wall; S-layer; Calcium-binding.
FT INIT MET 0
FT SEQUENCE 1025 AA; 98209 MW; AFC8B519820B1A5F CRC64;
SQ
Query Match 10.8%; Score 101.5; DB 1; Length 1025;
Best Local Similarity 28.0%; Pred. No. 0.6;
Matches 54; Conservative 20; Mismatches 82; Indels 37; Gaps 7;
QY 6 VAIINSLTIVTITASASADPSKDSKAQV-----SAEAGITGTYWYNOLGSGFIYTAGAGC 61
DB 298 VQAAVYALPTGVITGISIETMNTSGAAILTNTSSGVTGLTALNTYSGAAQTVTAGAGC 357
QY 62 ALTGYESAVGNASRYVLTGRD-----SAPATDSSGALGVTVMKNNYRNHNS 112
DB 358 NLTAI-----TAAQANNAVAVDGRANVTASTGVTSCTTTVGANSASAGTV-----SVSYANS 410
QY 113 ATTWSGOY-VGGAERINTQWLLTSGTTEANAMKSTLVGH-----TFTKVPSSA 162
DB 411 STTTGAIIVTGTGTA-----VTVAQTGNNAVNTLTQADVTVTGNSSSTVATVYQTGA 463
QY 163 SIDAARAKAGYNG 175
DB 464 ATGATVAGRVNG 476
RESULT 13
MASZ_MYCTU STANDARD; PRT; 741 AA.
AC Q50596;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROBABLE MALATE SYNTHASE G (EC 4.1.3.2).
GN GLCB OR RV1837C OR MYCYIA11.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;

```

```

RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RL      Nature 393:537-544(1998).
CC      -1- CATABOLIC ACTIVITY: L-MALATE + COA = ACETYL-COA + H(2)O +
CC      GLYOXALATE.
CC      -1- PATHWAY: SECOND STEP IN GLYOXALATE BYPASS, AN ALTERNATIVE TO THE
CC      TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).
CC      -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE MALATE SYNTHASE G FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC      -----
DR      EMBL: 278020; CAB01465.1; -.
DR      TUBERCULIST; RV1837C; -.
KW      Glyoxylate bypass; Tricarboxylic acid cycle; Lyase.
SQ      SEQUENCE 741 AA; 80403 MW; A92F54E0FE8B7C64 CRC64;

Query Match
Best local similarity 10.3%; Score 96.5; DB 1; Length 741;
Matches 56; Conservative 28; Mismatches 86; Indels 41; Gaps 11;

OY      9  IAVSLTVSITASASAD-----PSKSKAQSAAEAGITGTWYNOLGSTFIYTAGADALNG 65
DB      100  ITSGVDAEITTTAGPQVLVPMARFALNANAR-WGSLYDLVGTDVPIE-TDGAENG 157
OY      66  -YESAVGNAESRYVTGKYDAPATDGS-CTALGWT-----YAKNNRYNAHSATTW 116
DB      158  PYYNKVRGKVIAYARKFLDDSVLSSGSGFATGFTVDGQVLVALPKSTGLANPGOF 217
OY      117  SGOYVGAGARINTOWLTLN-----GTTEANMKSTLV--GHDTFKKVP 159
DB      218  AG-YTGAAS--PTSYLLNHGHLHITLIDPESQVGTIDRAGKVDILSATTTIMDFED 274
OY      160  SASASIDAKKA-----GVNNGNPDLAVQO 183
DB      275  SVAAYDAADKVLGYRNMGLINKGDLAAAVDK 305

RESULT 14
ID      Y309_MYCGE STANDARD; PRT; 1225 AA.
AC      P47551; O49317;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOTHETICAL LIPOPROTEIN MG309 PRECURSOR.
GN      MG309.
OS      Mycoplasma genitalium.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC      Mycoplasmataceae; Mycoplasma.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-ATCC 33530 / G-37;
RX      MEDLINE-96026346; PubMed=756993;
RA      Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA      Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA      Fritchman J.L., Weidman J.F., Small K.V., Sandhu S.M., Fuhmann J.L.,
RA      Nguyen D.T., Usterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
RA      Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Luetter T.S.,
RA      Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA      "The minimal gene complement of Mycoplasma genitalium.";
RL      Science 270:357-403(1995).
RN      [2]
RN      SEQUENCE OF 1138-1224 FROM N.A.
RN      PC STRAIN-ATCC 33530 / G-37;

```

```

RX      MEDLINE-94075230; PubMed-8253680;
RA      Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III.
RT      "A survey of the Mycoplasma genitalium genome by using random
RT      sequencing."
RL      J. Bacteriol. 175:7918-7930(1993).
CC      -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC      ANCHOR (POTENTIAL).
CC      -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC      -----
DR      EMBL: U39711; AAC71531.1; -.
DR      EMBL: U02200; AAD12488.1; -.
DR      TIGR: MG309; -.
DR      PROSITE: PS00013; PROKAR-LIPOPROTEIN; 1.
KW      Hypothetical protein; Lipoprotein; Membrane; Signal.
FT      SIGNAL 1 27
FT      CHAIN 28 1225
FT      LIPID 28 28
FT      CONFLICT 1185 1185
SQ      SEQUENCE 1225 AA; 138375 MW; C3E4BF5B4319B6E8 CRC64;

Query Match
Best local similarity 9.9%; Score 93; DB 1; Length 1225;
Matches 56; Conservative 22; Mismatches 78; Indels 72; Gaps 9;

OY      4  IYVAIAVSLTVSITASADPS-----KSKAQSAAEAGITGTWY----NOLGS 51
DB      12  LLTSSIAVSL--GIYAVACAPNSRTIENLFPSSAFDKNDGSIATVLYALENREGL 68
OY      52  TFIYTAGAGALGTYESAVGNAESRYVTGKYDAPATDGSCTALGWTYAKNNRYNANH 111
DB      69  TQYTLWRILAPVRKNFEENVDDIKRNLFTFTD-----TDNS-----FVNOEOLRN-- 116
OY      112  SATWMSGQYVGAGARINTOWLTLN-----GTTEANMKSTLV--GHDTFKKVP 159
DB      117  -----QYRGDIYVRLQDILDNITGNGQAN-WKLRDVANNKIVDEPINKLFTKNFEYVD 168
OY      157  -----VKPSASIDAKKAGVNNGNPLDAVQ 182
DB      169  KSYGVLTPLKGLIENOSNMNMIKIQAKFVDRKKRLRINNDAVYAIQ 216

RESULT 15
ID      STRF_ECOLI STANDARD; PRT; 1120 AA.
AC      P76072; P77560;
DT      01-OCT-2000 (Rel. 40, Created)
DT      01-OCT-2000 (Rel. 40, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC.
GN      STRF OR B1372.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-K12 / MG1655;
RX      MEDLINE-97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RA      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [2]

```



RP SEQUENCE FROM N.A.

RC STRAIN-K12:

RA MEDLINE-97251357; PubMed-9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,

RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,

RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.,

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE000234; AAC74454.1; ALT\_INIT.

DR EMBL: D90774; BAA14966.1; -

DR EMBL: D90775; BAA14975.1; -

DR ECOGENE: EG13370; STFR.

DR HSSP: P04002; 1MFA.

DR INTERPRO: IPR000122; -

KW Hypothetical protein; Fiber protein; Repeat.

SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

Query Match 9.8%; Score 91.5; DB 1; Length 1120;

Best Local Similarity 24.7%; Pred. No. 4.2;

Matches 42; Conservative 32; Mismatches 89; Indels 7; Gaps 4;

OY 10 AVSLTVSITASASADPPSKAQSNAEAGITGTWYNQLGFTFIYTAGADALGTYES 69

DB 219 ATASATITATASASADPPSKAQSNAEAGITGTWYNQLGFTFIYTAGADALGTYES 276

OY 70 AVGNASRYVLTGRYDSAPATDGGTALGTV-AMKNNYRNAHSATFTWSGQYVGAEART 128

DB 277 SETNARSSETTAAGQ--SASAAAGSKTAAASSASASTSAGQASASATTAAGKSAESAASSA 334

OY 129 NTQWLTSCTE--ANAKMSTLVGHDTFTYVKPSAASIDAAKAGVANGN 176

DB 335 STATRAGEATEQASAAARSASAKTSETNAKASETSAESSKTAASSAS 384

OY 129 NTQWLTSCTE--ANAKMSTLVGHDTFTYVKPSAASIDAAKAGVANGN 176

DB 335 STATRAGEATEQASAAARSASAKTSETNAKASETSAESSKTAASSAS 384

Search completed: April 6, 2001, 00:19:51

Job time: 155 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2001, 00:16:01 ; Search time 66.64 Seconds  
(without alignments)  
321.864 Million cell updates/sec

Title: US-09-589-870-2  
Perfect score: 936  
Sequence: 1 MRKIVVAIAVSLFTVSITA.....IDAAKKAGVNNGNPLDAVQQ 183

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL\_15:\*  
2: SP:Archea:\*  
3: SP:Bacteria:\*  
4: SP:human:\*  
5: SP\_Invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rudent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	114	12.2	529	5	Q25058 heliocidari
2	112	12.0	595	2	Q06353 escherichia
3	110.5	11.8	892	2	Q9KX38 moraxella c
4	105.5	11.3	1001	3	Q05164 saccharomyc
5	104	11.1	595	2	Q09507 escherichia
6	101	10.8	595	2	Q06351 escherichia
7	100.5	10.7	1026	2	Q09R12 caulobacter
8	100	10.7	565	2	Q09S42 escherichia
9	100	10.7	565	2	Q09S44 escherichia
10	99.5	10.6	1477	2	Q48028 haemophilus
11	98.5	10.5	832	2	Q54356 moraxella c
12	98	10.5	595	2	Q09S01 escherichia
13	97.5	10.4	729	2	Q9KH34 antarctic b
14	97	10.4	443	6	Q28243 canis fami
15	94.5	10.1	507	2	Q05056 streptomyce
16	94.5	10.1	381	2	Q09KX3 streptococ
17	94	10.0	570	2	Q59665 pseudomonas
18	94	10.0	967	3	Q08294 saccharomyc
19	92	9.8	865	2	Q43919 aeromonas c

20	92	9.8	1217	5	Q17240 bombyx mori
21	91.5	9.8	922	2	Q9L964 moraxella c
22	91.5	9.8	1058	2	P96141 tha1 tick t
23	91.5	9.8	1058	2	Q9ZT03 rickettsia
24	91.5	9.8	1060	2	P95642 rickettsia
25	91	9.7	231	2	Q9KXK3 streptococ
26	91	9.7	700	2	Q30678 xanthomonas
27	91	9.7	867	12	Q39782 equine herp
28	90.5	9.7	568	2	Q53789 shigella bo
29	90.5	9.7	568	2	Q53785 escherichia
30	90.5	9.7	1060	2	Q9LARI rickettsia
31	90.5	9.7	1060	2	Q9LARO rickettsia
32	90.5	9.7	2178	2	Q9KWR3 streptococ
33	90.5	9.7	2761	5	Q19522 caenorhabd
34	90	9.6	873	2	Q9L960 moraxella c
35	90	9.6	1183	2	Q09R07 cellulomona
36	90	9.6	13288	6	Q18758 sus scrofa
37	89	9.5	289	2	Q45977 caulobacter
38	89	9.5	866	12	Q39781 equine herp
39	88.5	9.5	1061	2	P95586 rickettsia
40	88.5	9.5	2021	2	Q52657 rickettsia
41	88	9.4	894	2	Q9L962 moraxella c
42	88	9.4	1060	2	Q9LAK7 rickettsia
43	87.5	9.3	470	5	Q9U6M5 glabodera t
44	87.5	9.3	574	2	Q9R9Y6 campylobact
45	87.5	9.3	1060	2	P95620 rickettsia

## ALIGNMENTS

RESULT 1  
ID Q25058 PRELIMINARY; PRT; 529 AA.  
AC Q25058;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE FIBROPELIN IA (FRAGMENT).  
OS Helicoidaris erythrogramma (Sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoldea; Euechinoldea; Echinacea; Echinolida; Echinometridae;  
OC Helicoidaris.  
OX NCBI\_TaxID:7634;  
RN  
RP  
RA  
RA BLISGROVE B.W.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L33861; AAA2995.1; -.  
DR HSSP; P08709; 1BF9.  
DR INTERPRO; IPR000083; -.  
DR INTERPRO; IPR000088; -.  
DR INTERPRO; IPR000152; -.  
DR INTERPRO; IPR000561; -.  
DR INTERPRO; IPR000742; -.  
DR INTERPRO; IPR001010; -.  
DR INTERPRO; IPR001438; -.  
DR INTERPRO; IPR001881; -.  
DR INTERPRO; IPR001947; -.  
DR INTERPRO; IPR002049; -.  
DR PEAM; PF00008; EGF\_10.  
DR PFAM; PF01382; Avidin; 1.  
DR PRINTS; PRO0010; EGFBLD.  
DR PRINTS; PRO0011; EGFFLAMININ.  
DR PRINTS; PRO0012; ENTPEI.  
DR PRINTS; PRO0286; CHARBYDOTOXIN.  
DR PRINTS; PRO0287; THIONIN.  
DR PRINTS; PRO0709; AVIDIN.  
DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_8.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_11.  
DR PROSITE; PS00577; AVIDIN; 1.  
DR PROSITE; PS01186; EGF\_2; 10.  
DR PROSITE; PS01187; EGF\_CA; 7.

KM Glycoprotein; EGF-like domain.  
 FT NON\_TER 1  
 SO SEQUENCE 529 AA; 55543 MW; DAAB958FCF9ACB5A CRC64;

Query Match 12.2%; Score 114; DB 5; Length 529;  
 Best Local Similarity 28.4%; Pred. No. 0.18;  
 Matches 40; Conservative 22; Mismatches 55; Indels 24; Gaps 7;

OY 34 VSAEAG---ITGWTYNOLGTFVTAGADGALITGTY---ESAVGNRESRYL---TGR 83  
 DB 396 VNCSEVGECDLEGWYNECNDQITITKSTGMLGDYMTAVEIAYVGAAPYVVGASNN 455  
 OY 84 YDSAPATDGSSTALGWTYVAMKNRYNNAHSATTWSGO-YVGEAARINTOMLLTSGT-TEA 141  
 DB 456 YDF-----PSGTFYVNRDNGM-----STTSWTAQCHLCDNEEVLTYTMINIMVDTQC 503  
 OY 142 NAWKSTVLGHDTFTKVPKPSAA 162  
 DB 504 DIKKNMVGDGDKMTRYSOSTA 524

RESULT 2  
 ID 006353 PRELIMINARY; PRT; 595 AA.  
 AC 006353;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE FLAGELLIN.  
 GN FLIC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BI 316-42 / SEROTYPE 09:K9:H12;  
 RX MEDLINE=93374833; PubMed=8366026;  
 RA Schoenhalz G., Whitfield C.;  
 RT "Comparative analysis of flagellin sequences from Escherichia coli strains possessing serologically distinct flagellar filaments with a shared complex surface pattern.";  
 RT J. Bacteriol. 175:5395-5402(1993).  
 CC -1 FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO FORM THE FILAMENTS OF BACTERIAL FLAGELLA.  
 CC EMBL: I07389; AAA3799.1; -.  
 DR INTERPRO: IPR001029; -.  
 DR INTERPRO: IPR001492; -.  
 DR PFAM: PF00669; Flagellin\_N; 1.  
 DR PFAM: PF00700; Flagellin\_C; 1.  
 DR PRINTS: PR00207; FLAGELLIN.  
 DR PRODOM: PD000316; -; 1.  
 KW Flagella.  
 SQ SEQUENCE 595 AA; 61044 MW; 88510CD561EF25F9 CRC64;

Query Match 12.0%; Score 112; DB 2; Length 595;  
 Best Local Similarity 26.1%; Pred. No. 0.29;  
 Matches 52; Conservative 30; Mismatches 87; Indels 30; Gaps 9;

OY 7 AATAVSLTTSITASADPSKSKAQAESAEGITGTWYNOLGST---FIYTAGDGL 63  
 DB 343 ASYTMGCTTYNFRGTGADAD-AATANAGVSFTDASKETVLANVATAKQKAVAAADGDTSA 401  
 OY 64 TGTYESAV-----GNAESRYVLGRYDSAPAT---DGSCTALG-PTVAMK--- 104  
 DB 402 ITTYKSGVOTYQAVFAAGGTASAKYADKADVSNATATYTDADGEMTTGTSTTKRSDA 461  
 OY 105 NNYRNHSTFTWSGOVVG--GAEARINTOMLLTS-CTTANAMKSTVLGHND---TFTKV 157  
 DB 462 NNGKTVVDSGTGTGKAPRVGAEEVYVYANGTLTTDATSGTVDKPLKALDEAISIDKF 521

OY 158 KPSASIDAAKKAGVNNGN 176  
 DB 522 RPSLGAIONRLDSAVTNLN 540

RESULT 3  
 ID 09KX38 PRELIMINARY; PRT; 892 AA.  
 AC 09KX38;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE USPAL.  
 GN USPAL.  
 OS Moraxella catarrhalis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae; Moraxella.  
 NCBI\_TaxID=480;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-O46E;  
 RX MEDLINE=20138164; PubMed=10671460;  
 RA Lafontaine E.R., Cope L.D., Aebi C., Latimer J.L., McCracken G.H. Jr., Hansen E.J.;  
 RT "The USPAL Protein and a Second Type of Usps2 Protein Mediate Adherence of Moraxella catarrhalis to Human Epithelial Cells in vitro.";  
 RT J. Bacteriol. 182:1364-1373(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O46E;  
 RA Maciver I., Latimer J.L., Cope L.D., Thomas S.E., Hansen E.J.;  
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U61725; AAF36416.1; -.  
 SQ SEQUENCE 892 AA; 93358 MW; 9D71A369672F44C7 CRC64;

Query Match 11.8%; Score 110.5; DB 2; Length 892;  
 Best Local Similarity 21.7%; Pred. No. 0.59;  
 Matches 43; Conservative 26; Mismatches 86; Indels 43; Gaps 6;

OY 17 SITASADPSKSKAQAESAEGITGTWYNOLGSTFIYTAGDGLT--GTYESAVGNA 74  
 DB 101 SYTGSSSWTAMGKESKSTIGGDTNDANGYTSITIGGYRAIGDSSITIGGYNOATG-- 158  
 OY 75 ESTYVYTGRTDAPATDGSCTALGWTVAMKNRY--NAHSATTWSGOVGAEARINTQ 131  
 DB 159 EKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSHNOATGEGSPFAGVENKANAN 218  
 OY 132 -----WLLTSGTTEANAK-STLVGHDTFTKV 158  
 DB 219 NAVAIGKNNFTIDGNSVAIGSNNTIDSGKQNVFLLGSSSTNTTAQSGSVLLGHNTAGK-- 276  
 OY 159 PSAASIDAAKKAGVNNGN 176  
 DB 277 -KATAVSSAKVNGTLIGN 293

RESULT 4  
 ID 005164 PRELIMINARY; PRT; 1001 AA.  
 AC 005164;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-AUG-1998 (TREMblrel. 07, Last annotation update)  
 DE AOB567, AOF1001, AOE110, AOE264 AND AOE130 GENES.  
 GN AOF1001.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomyces.  
 NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FY1679;

RA Gamo F.J., Lafuente M.J., Casamayor A., Aldea M., Casas C., Arto J.,  
 RA Herrero E., Gancedo C.,  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: X89715; CA61860.1;  
 SO SEQUENCE 1001 AA; 97673 MW; F11BC4522DCFBDAF CRC64;

Query Match 11.3%; Score 105.5; DB 3; Length 1001;  
 Best Local Similarity 24.7%; Pred. No. 1.7;  
 Matches 42; Conservative 33; Mismatches 64; Indels 31; Gaps 5;

OY 6 VAAIAVSLTIVSTASADPSKDSKAQVSAEAGITGTWYNOLGST---FIVAGADGAL 61  
 DB 63 VSSSLRELTSSSTEVSSSIAPSTSS---SEVSSITSSGSSVSGSSSTSSGSSSSSS 118  
 OY 62 ALTGTEASAVGNAESRYVLGTRDAPATDGGTALGWTYAMKNNRNHSAATTWGOYV 121  
 DB 119 SAATESGSSASGSSSAT-----EGSSVSGSSSTSI-----TSGSSATGSSSV 161  
 OY 122 GGAEARINTQWLTSCTTEANAMKSTLVGHDFTKVYKPSAASIDAAKKAG 171  
 DB 162 SGGSTSI-----TSGSSATGSSVSGSTSATGSSASGSSSATGSG 205

RESULT 5  
 ID 09S0T7 PRELIMINARY: PRT: 595 AA.  
 AC 09S0T7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE FLAGELLIN.  
 GN FLIC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BI 316-42;  
 RA Ohnishi K., Ishioka K., Matsuba T., Hareyama S.,  
 RT "Cloning of H antigen genes in E.coli serotypes and expression in  
 RT E.coli K-12."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB028475; BAA85084.1;  
 DR INTERPRO: IPR001029;  
 DR INTERPRO: IPR001492;  
 DR PFAM: PF00669; Flagellin\_N; 1.  
 DR PFAM: PF00700; Flagellin\_C; 1.  
 DR PRINTS: PR00207; FLAGELLIN.  
 SO SEQUENCE 595 AA; 61020 MW; 8B9DC3D6AC78C427 CRC64;

Query Match 11.1%; Score 104; DB 2; Length 595;  
 Best Local Similarity 25.6%; Pred. No. 1.2;  
 Matches 51; Conservative 30; Mismatches 88; Indels 30; Gaps 9;

OY 7 AAIASLITVSTASADPSKDSKAQVSAEAGITGTWYNOLGST---FIVAGADGAL 63  
 DB 343 ASYTMGTTYNFKTGADAD--AATANAGVSTPTDASKETVINKVATAKQGRAVAADGDTSA 401  
 OY 64 TGTYESAV-----GNAESRYVLGRYDSAPAT-----DGGSTALG-WTVAMK--- 104  
 DB 402 TITYKSGVQYQAVFAAGDGTASAKYADNDVSNATVTTDADGEMTTIGSYTTKYSIDA 461  
 OY 105 NNTNRNHSATTWGOYV--GAEARINTQWLTS--GTTEANAMKSTLVGHG----FTTKV 157  
 DB 462 NNGKVTYVDSTGKYPKGAIEVYSANGTLTTDATSEGTVTKDPLKALDEAISIDKF 521  
 OY 158 KPSAASIDAAKKAGVNNGN 176  
 DB 522 RSSLGAIONRLDSAVTNLN 540

RESULT 6  
 ID 006351 PRELIMINARY: PRT: 595 AA.  
 AC 006351;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE FLAGELLIN.  
 GN FLIC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SU 1242 / SEROTYPE 01:K2:H1;  
 RX MEDLINE=93374833; PUBMED=8366026;  
 RA Schoenhalz G., Whitfield C.;  
 RT "Comparative analysis of flagellin sequences from Escherichia coli  
 RT strains possessing serologically distinct flagellar filaments with a  
 RT shared complex surface pattern."  
 RL J. Bacteriol. 175:5395-5402(1993).  
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.  
 CC -1- SIMILARITY: NO OTHER BACTERIAL FLAGELLINS.  
 DR EMBL: L07387; AAA23797.1;  
 DR INTERPRO: IPR001029;  
 DR INTERPRO: IPR001492;  
 DR PFAM: PF00669; Flagellin\_N; 1.  
 DR PFAM: PF00700; Flagellin\_C; 1.  
 DR PRINTS: PR00207; FLAGELLIN.  
 DR PRODOM: PD000316; -; 1.  
 KW Flagella.  
 SO SEQUENCE 595 AA; 60923 MW; 9AAFB2E86884607A CRC64;

Query Match 10.8%; Score 101; DB 2; Length 595;  
 Best Local Similarity 25.6%; Pred. No. 2.1;  
 Matches 51; Conservative 29; Mismatches 89; Indels 30; Gaps 9;

OY 7 AAIASLITVSTASADPSKDSKAQVSAEAGITGTWYNOLGST---FIVAGADGAL 63  
 DB 343 ASYTMGTTYNFKTGADAG--AATANAGVSTPTDASKETVINKVATAKQGRAVAANDGDTSA 401  
 OY 64 TGTYESAV-----GNAESRYVLGRYDSAPAT-----DGGSTALG-WTVAMK--- 104  
 DB 402 TITYKSGVQYQAVFAAGDGTASAKYADNDVSNATVTTDADGEMTTIGSYTTKYSIDA 461  
 OY 105 NNTNRNHSATTWGOYV--GAEARINTQWLTS--GTTEANAMKSTLVGHG----FTTKV 157  
 DB 462 NNGKVTYVDSTGKYPKGAIEVYSANGTLTTDATSEGTVTKDPLKALDEAISIDKF 521  
 OY 158 KPSAASIDAAKKAGVNNGN 176  
 DB 522 RSSLGAIONRLDSAVTNLN 540

RESULT 7  
 ID 09RF12 PRELIMINARY: PRT: 1026 AA.  
 AC 09RF12;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE S-LAYER PROTEIN.  
 GN R5AA.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OX Caulobacter.  
 OX NCBI\_TaxID=76;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-JS3001;  
 RA Bingle W.H., Awram P.A., Nommellini J.F., Smit J.K.;  
 RT "The Secretion Signal of C. crescentus S-layer Protein is Located in  
 the C-terminal 82 Amino Acids of the Molecule.";  
 RL Submitted (OCT-1999) to the EMBL/Genbank/DBU databases.  
 DR EMBL; AF193063; AAF19365.1; -  
 DR INTERPRO: IPR001343; -  
 DR INTERPRO: IPR001952; -  
 DR PFAM: PF00353; hemolysincbind; 1.  
 DR PRINTS: PR00313; CAEMDNGRPT.  
 DR PROSITE: PS00123; ALKALINE\_PHOSPHATASE; UNKNOWN1.  
 DR SEQUENCE 1026 AA; 98132 MW; AD7DB818D7C528AC CRC64;

Query Match 10.7%; Score 100.5; DB 2; Length 1026;  
 Best Local Similarity 28.8%; Pred. No. 4.2;  
 Matches 55; Conservative 18; Mismatches 85; Indels 33; Gaps 7;

OY 6 VAAIAVSLTIVSITASADPSKDSKAQV-----SAAEAGITGWYNOLGSTFIYTAGADG 61  
 DB 299 VQAAAVTALPTGVTIGIEFMNTSGAAILTNSSGVTLGTLANTNTSGAAQTVTAGAGQ 358  
 OY 62 ALGTYESAV-----GNASRYVLTRYDSAPATDGSCTALGWTVMKNRYNAHSAT 114  
 DB 359 NLTATTAQAQANNVAVDGGANTVASTG-VTSGTTIVGANSASGTV---SVVANSST 413  
 OY 115 TWSGQY-VGGAERINTOMLITSCTEANAAMKSTLYGHD-----FTTKYKPSAS 164  
 DB 414 TTTGAIATVGTGTA-----VTVAQTAGNAVNNTLTQADVTYTGNSSTYAVVTQTAAAT 466  
 OY 165 DAAKAGVNNNG 175  
 DB 467 AGATVAGRVNG 477

RESULT 8  
 ID O9S4M2 PRELIMINARY; PRT; 560 AA.  
 AC O9S4M2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE FLAGELLIN (FRAGMENT).  
 GN FLIC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-E7A/68;  
 RX MEDLINE-99084952; PubMed-9864325;  
 RA Reid S.D., Selander R.K., Whittam T.S.;  
 RT "Sequence diversity of flagellin (flc) alleles in pathogenic  
 RT Escherichia coli.";  
 RL J. Bacteriol. 181:153-160(1999).  
 DR EMBL; AF128954; AAD28525.2; -  
 DR INTERPRO: IPR001023; -  
 DR INTERPRO: IPR001029; -  
 DR INTERPRO: IPR001492; -  
 DR PFAM; PF00669; Flagellin\_N; 1.  
 DR PFAM; PF00700; Flagellin\_C; 1.  
 DR PRINTS; PR00207; FLAGELLIN.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR PRINTS; PR00455; HTHETTR.  
 FT NON\_TER 1  
 FT SEQUENCE 560 AA; 57263 MW; 97DF005CE9E0ABB71 CRC64;

Query Match 10.7%; Score 100; DB 2; Length 560;  
 Best Local Similarity 25.1%; Pred. No. 2.4;

Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;  
 OY 7 AATAVSLTIVSITASADPSKDSKAQVSAEAGITGWYNOLGSI---FIYTAGADGAL 63  
 DB 324 ASVTMGSTTYNFKTAGDAD-AATANAGVSFTDPAKSEYVLNKVATAKQGRAAADGDTSA 382  
 OY 64 TGTYESAV-----GNASRYVLTRYDSAPAT-----DGSCTALG-WTVAMK--- 104  
 DB 383 TITTKSGVQTYQAVFAAGDGTASAKYADKADVSNATRTYTDADGEMTTIGSYTTKTSIDA 442  
 OY 105 NNYRNAHSATTWGQYVG--GAERINTOMLITS-GTTEANAAMKSTLYGHD---FTTKY 157  
 DB 443 NNGKVTYVDSGTGKTAAPKVGAEVYVSANGTLTTDATSEGTYVKDPLKALDEAIISSIDKF 502  
 OY 158 KPSASIDAAKKAQVNNGN 176  
 DB 503 RSSLGAIQNRLDSAVTNLN 521

RESULT 9  
 ID O9S4M4 PRELIMINARY; PRT; 565 AA.  
 AC O9S4M4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE FLAGELLIN (FRAGMENT).  
 GN FLIC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DEC 6A;  
 RX MEDLINE-99084952; PubMed-9864325;  
 RA Reid S.D., Selander R.K., Whittam T.S.;  
 RT "Sequence diversity of flagellin (flc) alleles in pathogenic  
 RT Escherichia coli.";  
 RL J. Bacteriol. 181:153-160(1999).  
 DR EMBL; AF128952; AAD28523.2; -  
 DR INTERPRO: IPR001023; -  
 DR INTERPRO: IPR001492; -  
 DR INTERPRO: IPR001029; -  
 DR INTERPRO: IPR001647; -  
 DR PFAM; PF00669; Flagellin\_N; 1.  
 DR PFAM; PF00700; Flagellin\_C; 1.  
 DR PRINTS; PR00207; FLAGELLIN.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR PRINTS; PR00455; HTHETTR.  
 FT NON\_TER 1  
 FT SEQUENCE 565 AA; 57857 MW; 6E9EE1A1AF72DC29 CRC64;

Query Match 10.7%; Score 100; DB 2; Length 565;  
 Best Local Similarity 25.1%; Pred. No. 2.4;  
 Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;

OY 7 AATAVSLTIVSITASADPSKDSKAQVSAEAGITGWYNOLGSI---FIYTAGADGAL 63  
 DB 328 ASVTMGSTTYNFKTAGDAD-AATANAGVSFTDPAKSEYVLNKVATAKQGRAAADGDTSA 386  
 OY 64 TGTYESAV-----GNASRYVLTRYDSAPAT-----DGSCTALG-WTVAMK--- 104  
 DB 387 TITTKSGVQTYQAVFAAGDGTASAKYADKADVSNATRTYTDADGEMTTIGSYTTKTSIDA 446  
 OY 105 NNYRNAHSATTWGQYVG--GAERINTOMLITS-GTTEANAAMKSTLYGHD---FTTKY 157  
 DB 447 NNGKVTYVDSGTGKTAAPKVGAEVYVSANGTLTTDATSEGTYVKDPLKALDEAIISSIDKF 506  
 OY 158 KPSASIDAAKKAQVNNGN 176







KW Hydrolase. 507 AA; 52681 MW; 5EAD34A9874C6A73 CRC64;  
SQ SEQUENCE

## Query Match

10.18; Score 94.5; DB 2; Length 507;

Best Local Similarity 27.6%; Pred. No. 5.8;

Matches 43; Conservative 16; Mismatches 52; Indels 45; Gaps 10;

QY 10 AVSLTIVS--ITASASADPS-----KDSKAQ-----VSAAEAGITGTYWYNQLGSTFIYTA 57  
          ::|||          ::|||          |||          |||          |||  
Db 355 SLSATPAGSTVTAGASATTTTKTAIVKSGTAQTVOLGASGVPAAGVTASF-----SPASVTA 409  
                                  -GTYESAV-----GNAESRYVLTGGRYDSAPATDQSG---TALGW 99  
QY 58 GADGALT-----GTYESAV-----GNAESRYVLTGGRYDSAPATDQSG---TALGW 99  
          ||          |||          |||          |||          |||  
Db 410 GGQSTLTLLATTGQAVSCTYSITYTGTCPSGSHSTTYALT-----VTGGNGNOCTAIVPW 462  
                                  ::|||          |||          |||          |||  
QY 100 T--VAMKNNYRNAHASATTWMSQ--YVGGAEARINTQW 132  
          :          :          :          :          :  
Db 463 NSGAIYTGGOQVSHGHHTWKAKKMWTTGGEPPGTGQW 498

Search completed: April 6, 2001, 00:18:32  
Job time: 151 sec

**THIS PAGE BLANK (USPTO)**



Quality:	122.50	Length:	172
Ratio:	1.250	Gaps:	10
Percent Similarity:	56.977	Percent Identity:	27.326

Align seg 1/1 to reverse of: AW632946 from: 1 to: 495

[illegible]

**MEDLINE** 99365275  
**COMMENT** Contact:

**COMMENT**

Contact: Britten RU  
Division of Biology  
California Institute of Technology  
1200 East California Blvd, Pasadena, CA 91125, USA  
Email: rbritten@caltech.edu.

FEATURES	Location/Qualifiers
source	1. .546

```
/organism="Strongylocentrotus purpuratus"  
/db_xref="taxon:7668"
```

```
/clone="g_C06"  
/clone_lib="Strongylocentrotus purpuratus early cleavage
```

```
stage embryo"
/dev_stage="early cleavage stage embryo"
```

BASE COUNT	145 a	138 c	132 g	131 t
ORIGIN				

```
alignment_scores:
```

Quality:	107.00	Length:	143
Ratio:	1.321	Gaps:	9
Percent Similarity:	56.643	Percent Identity:	27.972

alignment\_block;

Align seg 1/1 to: AF122254 from: 1 to: 546

[illegible]



Align seg 1/1 to reverse of: A1770339 from: 1 to: 560

```

10 AlaValSerLeuThrThrValSerIleThrAla..... 20
11 ||| ||| ..... |||||:|||||
507 GCGCGTGGGAGGTCGACTCGATCAGCGCCAGTTCTCGTGCACCTG 458
21 .....SerAlaSerAlaAspProSerIleAspSerIleValS 35
457 GCGCAGCCGCCCTCTCTCTCTCCCGCTGCAGCGCAACACCGGCCA 408
35 erAlaAlaGluAlaGlyIleThrGlyThrTyr...AsnGlnLeuGly 50
::|||::: ::|||::: |||||: |||||: ::|||:
407 CCGCGCGCCCGAGCTTGTCTGGCGGAGTGACACCAAGACCGCGCAGC 358
51 SerThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyTh 67
::: |||: |||: |||: |||: |||: |||: |||: |||:
357 GCGCGGTCAAGGCGTGGCGGAGCAGCGGCGGTCTCC..... 316
67 rGluserAlaValGlyAsnAlaGluSerArgTyr.....ValLeuThr 82
|||: |||: |||: |||: |||: |||: |||: |||: |||:
315 .....GCCAGCGCGTCTCAAGTACTGAGGATTAAGACGG 279
82 LysArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGly 98
::: |||: |||: |||: |||: |||: |||: |||: |||:
278 ACTCGCGGAGCGCGCGCGCTGTGACGACGACGCGGACCTTTTG... 232
99 TrpThrValAlaIleThrPlyAsnAsn..... 106
||| ||| |||: |||: |||: |||: |||: |||: |||: |||:
231 TGGACGGGGTGTGGCGCAGACGCTGTGCTTTGCCAGGTCTCTC 182
106 ..... 106
181 CTCACGATGCGAAGCTCAGCCCTTGACGCCGACGCCAGCTGCACGC 132
107 .....TyrArgAsnAlaHisSerAlaThrThrTyrSerGlyLeuTyr 120
|||: |||: |||: |||: |||: |||: |||: |||: |||:
131 GGTGTGTACGAGTGTGCGCCACGACCGCAACAGATCAGCGGGATGCC 82
121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnThrLeuThrSerG 137
::||| |||: |||: |||: |||: |||: |||: |||: |||:
81 GCCGGCGGTATAGCGCTCTCGAGCTCTTCATGGCTC..... 42
137 LysThrThrGluAlaAsnAlaThrPlySer 146
|||: |||: |||: |||: |||: |||: |||: |||: |||:
41 ..ACTGTGAACTCAAGATGTCCTCG 15

```

seq\_name: gb\_est79:C05809

seq\_documentation\_block:

LOCUS C05809 506 bp mRNA EST 16-OCT-1996  
DEFINITION C05809 Human pancreatic islet Homo sapiens cDNA clone hbc5182  
similar to elastase III B, mRNA sequence.

ACCESSION C05809.1 GI:1502585

VERSION C05809.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 506)

AUTHORS Takeda,J.

TITLE Human pancreatic islet ESTs

COMMENT Contact: Jun Takeda

INSTITUTE FOR MOLECULAR AND CELLULAR REGULATION, GUNMA UNIVERSITY

3-39-15 Showa-machi, Maebashi Gunma 371, Japan

Tel: 272-20-8856

Fax: 272-20-8896

Email: jtakeda@esb.gunma-u.ac.jp.

Location/Qualifiers

1..506

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="hbc5182"

/note="Vector: Lambda ZAP11; Site 1: Eco RI; Site 2: Xho I; mRNA was prepared from normal adult human islets. cDNA was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."

BASE COUNT 84 a 169 c 142 g 110 t 1 others

alignment\_scores: Quality: 94.50 Length: 153

Percent Similarity: 49.020 Percent Identity: 26.797

alignment\_block: US-09-589-870-2 x C05809 ..

Align seg 1/1 to: C05809 from: 1 to: 506

```

1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSe 17
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
13 ATGAGCTTGTGCTATGTTGCCAGGCTGTCTCAAACTCGTGGCGTCAAG 62
17 T.....IleThrAlaSerAlaS 23
63 CGATCCTCCCGCTCGCGCTCTCAAGACCTAGATTAACCTCAGGCT 112
23 erAlaAspProSerIleAspSerIleValSerAlaIleGluAla 39
::: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
113 ATGGCCACCTTCTCTCTGCGCTTCACGCCGCTGTCTCAAGTGAAGAT 162
40 GlyIleThrGlyThrTyr.....TyrAsnGlnLe 49
::: |||: |||: |||: |||: |||: |||: |||: |||: |||:
163 GCGGTCCCTACAGCTGTGCGGAGGTTCCCTGACGTAGAAAG 212
49 uGlySerThrPhe.....I 54
||| |||: |||: |||: |||: |||: |||: |||: |||: |||:
213 TGGAGCTTCAACACAGCTGTGGCGGTAGCCTCAGCCCCGACGTGG 262
54 LeValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrGlySerAla 70
::||| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
263 TTGTGACTGCCGCCACATCGATCTCTGACCTCTGACCTACAGTGTG 312
71 ValGlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSer.A 87
::|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
313 TTGGGCGAGTAGAC...CGTGTGTGAAGAGGCCCCCGACAGGTGAT 359
87 LysProAlaThrAspGly.....SerGlyThrAlaLeu 97
::|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
360 CCCATCAACCTGTGGGACCTTTGTGCATCACTGTGAACCGCTCGT 409
98 GlyTrpThrValAlaThrPlyAsnAsnTyrArgAsnAlaHisSerAlaTh 114
||| |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
410 GTGTGGCTGTGGCAGACATCGCCCTCATCAAGCTCTCAGCAGCGCC 459
114 TrpThrTyr 116
::|||
460 CAGCTGG 466

```

seq\_name: gb\_est36:AV432953

seq\_documentation\_block:

LOCUS AV432953 534 bp mRNA EST 23-AUG-2000  
DEFINITION AV432953 Porphyria yezoensis TU-1 Porphyria yezoensis cDNA clone  
PM022b09\_r 5', mRNA sequence.

ACCESSION AV432953.1 GI:8586178

VERSION AV432953.1

KEYWORDS EST.

SOURCE Porphyria yezoensis

ORGANISM Porphyria yezoensis

Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;

[illegible]

```

417 GCCATCGGT...GACGTCCGCTCCGCGTCGTCGCCCAATGCGCGAGGAC 463
164 e...AspAlaAlaLysLysAlaGlyValAlaAsnAsnGlyAsnProLeu 178
|||||
|||||
464 TGGTGAACGATCCGCGGTGCGCTCAAGACGACATGATGACCCCTTC 509
seq_name: gb_est47:AM677938

seq_documentation_block:
LOCUS      AM677938          476 bp      mRNA          EST          19-JUL-2000
DEFINITION WSI.12_H09.bl_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION  AM677938
VERSION    AM677938.1  GI:7551654
KEYWORDS   EST.
SOURCE     sorghum.
ORGANISM   Sorghum bicolor
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE  1 (bases 1 to 476)
AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudan,M. and Pratt
            L.H.
TITLE      An EST database from Sorghum: water-stressed plants
JOURNAL   Unpublished (2000)
COMMENT    Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: emmpratt@uga.edu
            Sequences have been trimmed to exclude polyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 405
            POLYA-No.

FEATURES
source     1..476
            /organism="Sorghum bicolor"
            /db_xref="taxon:4558"
            /clone_1lb="Water-stressed 1 (WS1)"
            /note="Organ: Mix of 5-week old plants on days 7 & 8 after
            water was withheld; Vector: Lambda Zap; Site_1: XhoI;
            Site_2: EcoRI; The library was made from poly-A RNA in the
            cloning vector lambda Zap II. Clones to be sequenced were
            prepared by mass excision."
BASE COUNT      102 a      141 c      173 g      60 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
8
```

```

60 AspglyalaleuthrGlyThrTyrGluseralAvalGlyAsnAlaGlu 76
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 TCGGCTCTCTTA...GGACCTCGGTCTCCGCCCGGAGCTGCTCTC 265
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 rArgTyValLeuthrGlyArgTyAspSerAlaProAlaThrAspGly 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
264 GCGGCGCGGCTCTCTTCGACACCGGAGCGAGCCCTCTCTGGGTG 215
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 ..SerGlyThrAlaLeuGlyTyrThrValAlaTrpLysAsnAsnTyrArg 108
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 TTTCCTCAACGCGCGCGCTGACAGCC.....TCTCCACCTACC 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 AsnAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
173 ACCGAGGCTCTCTCGACGGCGACGAGTCGGACCTGACCTCGCGGT 124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
120 rValGlyGlyAlaGluAlaArgIleAsnThrGlnTyrLeu 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 GGCCTATGGGTCA...GCTAGGCTG.....CTGTGGTTG 93

```

seq\_name: gb\_est70:BE355782

seq\_documentation\_block: 477 bp mRNA EST 20-JUL-2000  
 LOCUS BE355782  
 DEFINITION DGI\_11.A01.b1.A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA

ACCESSION BE355782  
 VERSION BE355782.1 GI:9296884  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 477)  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.  
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 /L.H.

TITLE An EST database from Sorghum: dark-grown seedlings  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: JEN REV  
 High quality sequence stop: 385  
 POLYA-No.

FEATURES  
 source Location/Qualifiers  
 1..477

/organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda Zap II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 96 a 146 c 174 g 61 t  
 ORIGIN

#### alignment\_scores:

Quality: 94.00 Length: 130  
 Ratio: 1.106 Gaps: 7  
 Percent Similarity: 65.385 Percent Identity: 31.538

#### alignment\_block:

US-09-589-870-2 x BE355782/rev ..

Align seg 1/1 to reverse of: BE355782 from: 1 to: 477

```

10 AlAvalSerLeuthrThrValSerIleThrAlaSerAlaAspPr 26
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
437 TCTGCTCTCTTGACCTCC...TCCGTCTCTGCTCCGCCGCGCTCGGC 391
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
26 oSerTyAspSerIysAlaGluValSerAlaAlaGluAlaGlyIleThr 43
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 GCGCTCTGCTCTTGCTGTCTGCTCTGCTCGCCGCGGCTCACACG 341
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 lYrThrTrpTyAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 CCTGACCTGACCTCTGTGGTCTCAGCTCGGACGAGACAGCTTCGCC 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 AspglyalaleuthrGlyThrTyrGluseralAvalGlyAsnAlaGlu 76
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 TCCGCTCTCTTA...GGACCTCGGTCTCCGCCCGGAGCTGCTGTC 244
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 rArgTyValLeuthrGlyArgTyAspSerAlaProAlaThrAspGly 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
243 GCGGCGCGGCTCTCTTCGACACCGGAGCGAGCCCTCTCTTGGGTG 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 ..SerGlyThrAlaLeuGlyTyrThrValAlaTrpLysAsnAsnTyrArg 108
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 TTTCCTCAACGCGCGCGCTGACAGCC.....TCTCCACCTACC 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 AsnAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 ACCGAGGCTCTCTCGACGGCGACGAGTCGGACCTGACCTCGCGGT 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
120 rValGlyGlyAlaGluAlaArgIleAsnThrGlnTyrLeu 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 GGCCTATGGGTCA...GCTAGGCTG.....CTGTGGTTG 72

```

seq\_name: gb\_est70:BE357728

seq\_documentation\_block: 477 bp mRNA EST 20-JUL-2000  
 LOCUS BE357728  
 DEFINITION DGI\_22.C07.b1.A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA

ACCESSION BE357728  
 VERSION BE357728.1 GI:9299285  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 477)  
 Magnoliophyta; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.  
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 /L.H.

TITLE An EST database from Sorghum: dark-grown seedlings  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: JEN REV  
 High quality sequence stop: 467  
 POLYA-No.

FEATURES  
 source Location/Qualifiers  
 1..477

/organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda Zap II.



BASE COUNT 98 a 144 c 174 g 61 t  
 ORIGIN

## alignment\_scores:

Quality: 94.00 Length: 130  
 Ratio: 1.106 Gaps: 7  
 Percent Similarity: 65.385 Percent Identity: 31.538

## alignment\_block:

US-09-589-870-2 x BE357728/rev ..

Align seg 1/1 to reverse of: BE357728 from: 1 to: 477

```

10 Alvalserleuthrthrvalserlethralsaseralasaprr 26
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 TCTGCTCTGCTGACCTCC...TCCGCTCTGCTGCGCGCGCTCGGC 406
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 oserlyaspserysalaglnvalseralalaglualglyliethrg 43
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 CGGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 lthrtrptryasnnglnleuglyserthrphelvalthrallaglyala 59
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 CCTGACCTGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 Aspqlyalaleuthrthrtrgluseralalvalglyasnalagluse 76
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 TCCGCGCTCCTTA...GCGACCTCGCTCCGCCCGCGAGCTGCTGTC 259
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 rargtyvalleuthrthrlyargtyraspseralaproalathrasply. 92
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 GGGCGCGGGCTCTCTGACACCGACGAGCGCTCTCTGCTGCTG 209
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 ..SerglythrAlaleuGlyTrpThrValAlatrPluYsAsnAsnTyArg 108
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208 TTTCCTCAACGGCGCGCGCTGACAGCC.....TCTCCACTACAC 168
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 AsnAlahlsSerAlatr.....ThrTrpSerGlyGlnTy 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ACCGAGCCTCTCTGACGCGACGCGAGCTGCGGACCTGCGCGGT 118
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 rValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpleu 133
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 GGCATTGGTGCA...GCTAGGGTG.....CTGTGGTTG 87

```

seq\_name: gb\_est70:BE357007

## seq\_documentation\_block:

LOCUS BE357007 493 bp mRNA EST 20-JUL-2000  
 DEFINITION DGI\_146.D05.b1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 sequence.

ACCESSION BE357007  
 VERSION BE357007.1 GI:9298564

## KEYWORDS

EST.

## SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@atcuga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence is 20.  
 seq primer: JEN REV  
 High quality sequence stop: 402  
 POLYA-No.

FEATURES  
 source Location/Qualifiers

1..493  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DGI)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 104 a 147 c 179 g 63 t  
 ORIGIN

## alignment\_scores:

Quality: 94.00 Length: 130  
 Ratio: 1.106 Gaps: 7  
 Percent Similarity: 65.385 Percent Identity: 31.538

## alignment\_block:

US-09-589-870-2 x BE357007/rev ..

Align seg 1/1 to reverse of: BE357007 from: 1 to: 493

```

10 Alvalserleuthrthrvalserlethralsaseralasaprr 26
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 TCTGCTCTGCTGACCTCC...TCCGCTCTGCTGCGCGCGCTCGGC 413
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 oserlyaspserysalaglnvalseralalaglualglyliethrg 43
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 CGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 lthrtrptryasnnglnleuglyserthrphelvalthrallaglyala 59
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 CCTGACCTGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 313
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 Aspqlyalaleuthrthrtrgluseralalvalglyasnalagluse 76
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 TCCGCGCTCCTTA...GCGACCTCGGTCCGCCCGCGAGCTGCTGTC 266
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 rargtyvalleuthrthrlyargtyraspseralaproalathrasply. 92
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 GGCAGCGGGCTCTCTGACACCGACGAGCGCTCTCTGCTGCTG 216
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 ..SerglythrAlaleuGlyTrpThrValAlatrPluYsAsnAsnTyArg 108
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 TTTCCTCAACGGCGCGCGCTGACAGCC.....TCTCCACTACAC 175
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 AsnAlahlsSerAlatr.....ThrTrpSerGlyGlnTy 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 ACCGAGCCTCTCTGACGCGACGCGAGCTGCGGACCTGCGCGGT 125
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 rValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpleu 133
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 GGCATTGGTGCA...GCTAGGGTG.....CTGTGGTTG 94

```

seq\_name: gb\_est70:BE359336

## seq\_documentation\_block:

LOCUS BE359336 511 bp mRNA EST 20-JUL-2000  
 DEFINITION DGI\_40.E08.b1\_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 sequence.

ACCESSION BE359336  
 VERSION BE359336.1 GI:9300893

## KEYWORDS

EST.

## SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@atcuga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions

```

REFERENCE      1 (bases 1 to 511)
AUTHORS        Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
                L.H
TITLE          An EST database from Sorghum: dark-grown seedlings
JOURNAL        Unpublished (2000)
COMMENT        Contact: Cordonnier-Pratt MM
                Department of Botany
                The University of Georgia
                Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
                Tel: 706 542 1860
                Fax: 706 542 1805
                Email: mmp@pratt@uga.edu
                Sequences have been trimmed to exclude PolyA, vector and regions
                below phred quality 16. The threshold for highest quality sequence
                is 20.
                Seq primer: JEN REV
                High quality sequence stop: 490
                POLYA-No.

FEATURES
  source        1. 511
                Location/Qualifiers
                /organism="Sorghum bicolor"
                /db_xref="taxon:4558"
                /clone_lib="Dark Grown 1 (DG1)"
                /note="Organ: 5-day-old dark-grown seedlings; Vector:
                Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                made from poly-A RNA in the cloning vector lambda Zap II.
                Clones to be sequenced were prepared by mass excision."

BASE COUNT     107 a 154 c 166 g 64 t

ORIGIN
alignment_scores:
  quality:      94.00      length:      130
  ratio:         1.106      gaps:         7
  percent similarity: 65.385      percent identity: 31.538

alignment_block:
  US-09-589-870-2 x BE359336/rev ..
  Align seg 1/1 to reverse of: BE359336 from: 1 to: 511

10 AlavAlSerLeuThrThrValSerIleThrAlaSerAlaSerAlaAspPr 26
   ::::::::::::::::::::|::::::::::::::::::|
460 TCTGCTCTGCTTGGCTCTGCTCTGCTCTGCTCGCGCGCTCAGCCACGG 414
   ::::::::::::::::::::|::::::::::::::::::|
26 oSerLyAspSerLyAlaGlnValSerAlaAlaGlnAlaGlyIleThrg 43
   ::::::::::::::::::::|::::::::::::::::::|
413 CGCCTCTGCTTGGCTCTGCTCTGCTCTGCTCGCGCGCTCAGCCACGG 364
   ::::::::::::::::::::|::::::::::::::::::|
43 lYhTrpTyraSngInLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::::::::::::::::::|::::::::::::::::::|
363 CCTGACCTCGACCTCCTGCTCAGTCTCGGACGAGACAGACTCTCGGCC 314
   ::::::::::::::::::::|::::::::::::::::::|
60 AspGlyAlaLeuThrGlyThrTygluSerAlaValAlaGlySnaIgluSe 76
   ::::::::::::::::::::|::::::::::::::::::|
313 TCGCGCTCTCTTA...GCGACCTCGCTCTCGCGCGCTCAGCTCTGCTC 267
   ::::::::::::::::::::|::::::::::::::::::|
76 rArgTyValLeuThrGlyArgTyraSperAlaProAlaThrAspGly. 92
   ::::::::::::::::::::|::::::::::::::::::|
266 GCGCGCGCGGCTCTCTTCGACACCGGACGAGAGGCTCTCTCTGGGTG 217
   ::::::::::::::::::::|::::::::::::::::::|
93 ..SerGlyThrAlaLeuGlyTyTrpThrValAlaTrpIlySAsnAsnTyra 108
   ::::::::::::::::::::|::::::::::::::::::|
216 TTTCCTCAACGGCGCGCGCTCGACACC.....TCTCCACATCAC 176
   ::::::::::::::::::::|::::::::::::::::::|
109 AsnAlaHisSerAlaThr.....ThrTrpSerGlyGlnTy 120
   ::::::::::::::::::::|::::::::::::::::::|
175 ACCGAGACCTCTCTCGACGCGACGCGACGCTCGGACCTCTCGCGGT 126
   ::::::::::::::::::::|::::::::::::::::::|
120 rValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeu 133
   ::::::::::::::::::::|::::::::::::::::::|
125 GCGCATTTGGTCA...GCTAGGGTG.....CTGTGGTTG 95

```

```

seq_name: gb_est70:BE361106
seq_documentation_block:
  LOCUS      BE361106      524 bp      mRNA      EST      20-JUL-2000
  DEFINITION DGI.69.F08.b1.A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
  sequence.
  ACCESSION  BE361106
  VERSION    BE361106.1 GI:9302663
  KEYWORDS   EST.
  SOURCE     sorghum.
  ORGANISM   Sorghum bicolor
                Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
                Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE   1 (bases 1 to 524)
AUTHORS     Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
                L.H.
TITLE       An EST database from Sorghum: dark-grown seedlings
JOURNAL     Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
                Department of Botany
                The University of Georgia
                Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
                Tel: 706 542 1860
                Fax: 706 542 1805
                Email: mmp@pratt@uga.edu
                Sequences have been trimmed to exclude PolyA, vector and regions
                below phred quality 16. The threshold for highest quality sequence
                is 20.
                Seq primer: JEN REV
                High quality sequence stop: 520
                POLYA-No.

FEATURES
  source        1. 524
                Location/Qualifiers
                /organism="Sorghum bicolor"
                /db_xref="taxon:4558"
                /clone_lib="Dark Grown 1 (DG1)"
                /note="Organ: 5-day-old dark-grown seedlings; Vector:
                Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                made from poly-A RNA in the cloning vector lambda Zap II.
                Clones to be sequenced were prepared by mass excision."

BASE COUNT     107 a 156 c 195 g 66 t

ORIGIN
alignment_scores:
  quality:      94.00      length:      130
  ratio:         1.106      gaps:         7
  percent similarity: 65.385      percent identity: 31.538

alignment_block:
  US-09-589-870-2 x BE361106/rev ..
  Align seg 1/1 to reverse of: BE361106 from: 1 to: 524

10 AlavAlSerLeuThrThrValSerIleThrAlaSerAlaSerAlaAspPr 26
   ::::::::::::::::::::|::::::::::::::::::|
456 TCGCTTCTCTGACCTCC...TCCGCTCTGCTCTCGCGCGCGCTCTCGG 410
   ::::::::::::::::::::|::::::::::::::::::|
26 oSerLyAspSerLyAlaGlnValSerAlaAlaGlnAlaGlyIleThrg 43
   ::::::::::::::::::::|::::::::::::::::::|
409 CGCCTCTGCTTGGCTCTGCTCTGCTCTGCTCGCGCGCTCAGCCACGG 360
   ::::::::::::::::::::|::::::::::::::::::|
43 lYhTrpTyraSngInLeuGlySerThrPheIleValThrAlaGlyAla 59
   ::::::::::::::::::::|::::::::::::::::::|
359 CCTGACCTCGACCTCCTGCTCAGTCTCGGACGAGAACAGCTTCGCCG 310
   ::::::::::::::::::::|::::::::::::::::::|
60 AspGlyAlaLeuThrGlyThrTygluSerAlaValAlaGlySnaIgluSe 76
   ::::::::::::::::::::|::::::::::::::::::|
309 TCGGCTCTCTTA...GCGACCTCGCTCTCGCGCGCGGAGCTGCTGCTC 263
   ::::::::::::::::::::|::::::::::::::::::|
76 rArgTyValLeuThrGlyArgTyraSperAlaProAlaThrAspGly. 92
   ::::::::::::::::::::|::::::::::::::::::|
262 GCGCGCGCGGCTCTCTTCGACACCGGACGAGGCGCTCTCTCTGGGTG 213

```

```

93 ..SerglyThralaenglYtrPthrValAlatrPlyAsnAsnTyArg 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212 TTCTCTACAGCGCGCGCGCTGACAGCC.....TCCTCCACTACC 172
109 AsnAlaHisSerAlaThr.....ThrTrpSerglyGlnTy 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 ACCGAGCCTCTCTCGACGCGCAGCACTGCGGACCTGCGCGCGT 122
120 rValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeu 133
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GGCCATTGCTGCA...GCTAGGCTG.....CTGCGTTG 91

seq_name: gb_est70:BE357620

seq_documentation_block: 538 bp mRNA EST 20-JUL-2000
LOCUS BE357620
DEFINITION DGI_21.C08.b1.A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE357620
VERSION BE357620.1 GI:9299177
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
1 (bases 1 to 538)
Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
CONTACT: Cordonier-Pratt MM
DEPARTMENT OF Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@pratt.uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 534
POLYA-No.

FEATURES
source 1..538
location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced by mass excision."

BASE COUNT 109 a 161 c 203 g 65 t
ORIGIN

alignment_scores:
Quality: 94.00 Length: 130
Ratio: 1.106 Gaps: 7
Percent Similarity: 65.385 Percent Identity: 31.538

alignment_block:
US-09-589-870-2 x BE357620/rev ..
Align seq 1/1 to reverse of: BE357620 from: 1 to: 538

10 AlAtValSerLeuThrThValSerIleThrAlaSerAlaAspPr 26
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 TCTGCTTCCTTGACCTCC...TCGTCCTCTGCGCGCGCGCTCGGC 406
26 oSerLyAspSerLyAlaGluAlaArgIleAsnThrGlnTrpG 43
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

405 CGCCTCTGCTTCCTTGCTGCTGCTGCTGCTGCGCGCGCTCAGCCAGC 356
43 LYThrTrpTrpAsnGlnIleuGlySerThrPheIleValThrAlaGlyAla 59
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 CCTGACCTCGACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306
60 AspGlyAlaLeuThrGlyThrTyrcIuSerAlaValGlyAsnAlaGlu 76
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305 TCCGCTCTCTTA...GCGACCTCGCTCTCCGCGCGCGAGCTGCTCTC 259
76 rArgTyValLeuThrGlyArgTyTrpSerAlaProAlaThrAspGly. 92
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 GCGCGCGCGCTCTCTGACACCGAGAGCGCTCTGCTGCTGCTGCTG 209
93 ..SerglyThralaenglYtrPthrValAlatrPlyAsnAsnTyArg 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 TTCTCTACAGCGCGCGCTGACAGCC.....TCCTCCACTACC 168
109 AsnAlaHisSerAlaThr.....ThrTrpSerglyGlnTy 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 ACCGAGCCTCTCTCGACGCGCAGCACTGCGGACCTGCGCGCGT 118
120 rValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeu 133
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 GGCCATTGCTGCA...GCTAGGCTG.....CTGCGTTG 87

seq_name: gb_est70:BE357934

seq_documentation_block: 558 bp mRNA EST 20-JUL-2000
LOCUS BE357934
DEFINITION DGI_23.D10.b1.A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE357934
VERSION BE357934.1 GI:9299491
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
1 (bases 1 to 558)
Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
CONTACT: Cordonier-Pratt MM
DEPARTMENT OF Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@pratt.uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 528
POLYA-No.

FEATURES
source 1..558
location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced by mass excision."

BASE COUNT 115 a 169 c 208 g 66 t
ORIGIN

alignment_scores:
Quality: 94.00 Length: 130
Ratio: 1.106 Gaps: 7

```

